Display
D
UNIVERSITY OF MICHIGAN
JUL 23 1957
MATH. ECON. LIBRARY

CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. IX - NO. 3
1957

Groupoïdes automorphes par le groupe cyclique	A. Sade	321
Some generalizations of Burnside's theorem	N. A. Wiegmann	336
Prime power representations of finite linear groups II	Robert Steinberg	347
Certain classes of ideals in polynomial rings	N. H. McCoy	352
On binomial coefficient residues	J. B. Roberts	363
Combinatorial properties of matrices of zeros and ones	H. J. Ryser	371
A class of non-Desarguesian projective planes	D. R. Hughes	378
Transitivities in projective planes	T. G. Ostrom	389
Dedekind completeness and a fixed-point theorem	E. S. Wolk	400
Cyclically ordered sets	M. I. Aissen	406
Certain groups of orthonormal step functions	J. J. Price	413
The closed ideals in an algebra of analytic functions	Walter Rudin	426
Representation of Banach algebras with an involution	J. A. Schatz	435
On countably paracompact normal spaces	M. J. Mansfield	443
Games on a compact set	J. E. L. Peck and A. L. Dulmage	450
On the inversion of the Gauss transformation	P. G. Rooney	459

Published for
THE CANADIAN MATHEMATICAL CONGRESS
by the
University of Toronto Press

EDITORIAL BOARD

H. S. M. Coxeter, A. Gauthier, R. D. James, R. L. Jeffery,
G. de B. Robinson, H. Zassenhaus

with the co-operation of

H. Behnke, R. Brauer, D. B. DeLury, G. F. D. Duff, I. Halperin,
W. K. Hayman, J. Leray, S. MacLane, P. Scherk, B. Segre,
J. L. Synge, W. J. Webber

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Everything possible should be done to lighten the task of the reader; the notation and reference system should be carefully thought out. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $8.00. This is reduced to $4.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of Alberta	Assumption University
University of British Columbia	Carleton College
Dalhousie University	Ecole Polytechnique
Université Laval	Loyola College
University of Manitoba	McGill University
McMaster University	Université de Montréal
Queen's University	Royal Military College
St. Mary's University	University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

GROUPOÏDES AUTOMORPHES PAR LE GROUPE CYCLIQUE

A. SADE

I. Introduction

1. *Définitions.* Nous appellerons *groupoïde* un ensemble non vide, G, muni d'une loi $(\times)$ faisant correspondre à tout couple ordonné $x, y \in G$, *au plus* un élément z de G, appelé produit de x par y, et satisfaisant à la loi d'homogénéité **(2)**. Si E est l'ensemble des z la loi $(\times)$ définit une *application de E sur G*. On dira que G est *incomplet* si $E \subset GG$ et *ordinaire* si $E = GG$.

Un *semi-groupe* **(5)** est un groupoïde associatif.

Un *quasigroupe*, ordinaire ou incomplet, est un groupoïde dont les éléments ont *au plus* un quotient à droite et un quotient à gauche. Le mot "diviseur" sera entendu au sens de "sous-quasigroupe".

Soit Q un groupe additif d'entiers, ou d'entiers mod. n, et T un sous-groupe de Q. Soit $(\times)$ une loi binaire des éléments de Q telle que:

(1) Q soit un groupoïde,

(2) l'application $x \to x + t$, $t \in T$, soit un automorphisme de Q. La structure ainsi définie est appelée un *groupoïde automorphe par le groupe cyclique T* et est désignée par $Q(\times)$. Si $Q = T$ on dira que $Q(\times)$ est *automorphe par le groupe cyclique*.

2. *Exemple* I. L'ensemble des nombres rationnels muni de la loi:

$$x \times y = \tfrac{1}{2}(x + y) + C$$

3. *Exemple* II. Le quasigroupe:

$$x \times y = -(x - y - 1)^4 + 4(x - y - 1) + y + 5 \pmod 7$$

4. *Un groupoïde $G(\times)$ automorphe par le groupe cyclique est entièrement défini par la fonction $f(x) = x \times 0$*, ou par l'application

$$S: \quad x \to xS = f(x),$$

d'un certain sous-ensemble de G, dans G; la fonction $f(x)$ étant définie pour tout élément de ce sous-ensemble.

5. La condition exprimant qu'un groupoïde, automorphe par le groupe cyclique, est un semi-groupe n'est pas simple. Bornons-nous à ce résultat: *Il n'existe aucun groupe automorphe par le groupe cyclique.*

Reçu le 7 avril, 1956.

6. *Pour qu'un groupoïde $G(\times)$, automorphe par le groupe cyclique, soit un quasigroupe il faut et il suffit que la fonction $x \times 0 = f(x)$ satisfasse aux deux conditions*:

$$(3) \qquad \begin{cases} \text{(i)} \ f(x) = f(y) \ \rightleftarrows \ x = y \\ \text{(ii)} \ x - f(x) = y - f(y) \ \rightleftarrows \ x = y \end{cases}$$

(mod. n si G est fini et d'ordre n.)

La première exprime la loi de cancellation à droite (N° 1) pour $y = 0$. Montrons qu'elle est obéie par deux facteurs quelconques. L'égalité:

$$x \times a = y \times a,$$

s'écrit, en vertu de l'automorphisme:

$$(x - a) \times 0 + a = (y - a) \times 0 + a$$
$$f(x - a) = f(y - a)$$

Donc (i)
$$x = y.$$

(mod. n si G est fini.)

La seconde exprime la loi de cancellation à gauche.

Les réciproques sont immédiates.

7. Les quasigroupes infinis, automorphes par le groupe cyclique, ont peu de propriétés. Toute fonction $f(x)$ ayant une dérivée toujours négative, ou toujours plus grande que un, ou toujours comprise entre 0 et 1, fournit une solution.

8. *Tout quasigroupe ordinaire fini, automorphe par le groupe cyclique, est d'ordre impair.*

Le quasigroupe étant ordinaire, $f(x)$ décrit toutes les valeurs de x; donc, en désignant par n l'ordre du quasigroupe:

$$\Sigma f(x) = 0 + 1 + 2 + \ldots + (n - 1) = S \qquad \text{(mod. } n\text{)}.$$

D'après (ii), N° 6, la fonction $x - f(x)$ décrit ce même ensemble, donc:

$$\Sigma [x - f(x)] = S = S - S \qquad \text{(mod. } n\text{)},$$
$$S = n(n - 1)/2 \equiv 0 \qquad \text{(mod. } n\text{)}$$

$n = 2M + 1$.

Mais il existe des quasigroupes d'ordre pair *incomplets*, automorphes par le groupe cyclique.

Exemples: $x = 0\ 1\ 2\ 3$; et $x = 0\ 1\ 2\ 3\ 4\ 5$
$f(x) = 2\ 0 - 3$ $\qquad$ $f(x) = 5\ 4\ 2 - 0\ 3$.

Le quasigroupe incomplet du 6$^{\text{ième}}$ ordre: $x = 0\,1\,2\,3\,4\,5$; $x \times 0 = 3\,0\,1\,4\,2\,5$; $x \times 1 = 1\,5\,2 - 0 -$; $x \times 2 = 4\,1\,5\,2\,3\,0$; $x \times 3 = 2 - 3\,1\,4 -$; $x \times 4 = 5\,2\,0\,3\,1\,4$; $x \times 5 = 0 - 4 - 5\,3$, est automorphe par le sous-groupe du 3^{e} ordre du groupe cyclique:

$$T = 024.135;\ T^2;\ T^3 = 1.$$

Il existe un quasigroupe ordinaire, automorphe par $x \to x + i$, $(i = 0, 3, 6)$ et défini par:

$$x = 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ ; \qquad x \times 1 = 2\ 6\ 1\ 7\ 3\ 0\ 4\ 8\ 5$$
$$f(x) = 6\ 1\ 5\ 3\ 2\ 7\ 0\ 4\ 8\ ; \qquad x \times 2 = 8\ 5\ 3\ 4\ 7\ 1\ 2\ 0\ 6.$$

II. Transformations préservant l'invariance par le groupe cyclique.

9. D'après le N° 4 on peut identifier tout quasigroupe automorphe par le groupe cyclique avec la fonction correspondante:

$$x \times 0 = f(x)$$

et parler du quasigroupe $f(x)$ mod. n. On peut même le représenter par la substitution $x \to f(x)$ mise sous la forme de produit de cycles.

10. *Si $f(x)$ définit un quasigroupe automorphe par le groupe cyclique, les fonctions*:

a:	$x - f(x)$,	
b:	$-f(-x)$,	
$ab = ba$:	$x + f(-x)$	(conjoint (**3** p. 60, N° 75)),
c:	$f^{-1}(x)$	(réciproque),
D_h:	$f(x + h)$,	
E_h:	$f(x) + h$,	

jouissent de la même propriété.

Soit $f(x)$ une fonction satisfaisant aux conditions (3) du N° 6. Alors,

(*a*) $$g(x) = x - f(x)$$

sera encore une solution, car

(i) $$g(x) = g(y) \rightleftarrows x - f(x) = y - f(x) \rightleftarrows x = y,$$

puisque $f(x)$ est une solution.

(ii) $$x - g(x) = f(x),$$

donc

$$x - g(x) = y - g(y) \rightleftarrows f(x) = f(y) \rightleftarrows x = y.$$

Les autres parties s'établissent de manière analogue.

Si $f(x)$ est d'ordre n, les n opérations E_h, $(h = 0, 1, \ldots, n-1)$ forment un groupe cyclique car, si l'on désigne chaque opération par la valeur correspondante de h, ce groupe est isomorphe au groupe additif des restes, mod. n.

11. *Groupes formés par les opérations précédentes.* Considérons l'ensemble F de toutes les fonctions $f(x)$ solutions des conditions (3) pour une valeur donnée de n. Chacune de ces opérations a, b, c, D, E fait correspondre à chaque $f(x)$ une $f(x)$ et une seule et réciproquement. Ce sont des transformations de F. Elles forment un groupe, sous-groupe du groupe total (**8**, p. 3).

Les opérations a, b et les opérations b, c engendrent deux *groupes carrés*. En effet, $ab = ba$ (N° 10), et $bc = cb$, car l'une comme l'autre transforme $x \to f(x)$ en $x \to -f^{-1}(-x)$.

L'opération $ca = (ac)^{-1}$ *est du* $6^{\text{ème}}$ *ordre et* $(ca)^3 = b$.
Les opérations a, b, c *engendrent un groupe diédral du* 12^{me} *ordre*, défini par:

$$ab = ba,\ bc = cb,\ a^2 = b^2 = c^2 = 1,\ (ac)^3 = (ca)^3 = b.$$

12. *Les opérations* E_h *et* D_h *sont transformées l'une de l'autre par* c.

$$c^{-1}E_hc = D_h.$$

13. *D et E sont des isotopies* **(1)**. Si $D_h = E_k$, en posant $f(x + h) = z$ et $f(x) = y$, on aura:

$$f(x + h)E_k = z + k$$

et

$$f(x + h)D_h = y,$$

d'où:

$$y = z + k,$$

ou:

$$f(x) = f(x + h) + k,$$

et enfin

$$f(x) = ax + b \quad (ah + k \equiv 0, \text{ mod. } n).$$

Pour que $f(x)$ et $x - f(x)$ décrivent toutes les valeurs de x il faut de plus que a et $a - 1$ soient premiers avec n. Ainsi:

Les quasigroupes $f(x)$ *pour lesquels toute opération D est aussi une opération E sont* $f(x) = ax + b$, *où* a *et* $a - 1$ *sont premiers avec* n.

Les opérations (a), (b), (c) du N° 10 laissent cette propriété invariante.
Si $a = (n + 1)/2$, on aura un quasigroupe abélien.

14. Voici les lois de composition de tous les quasigroupes ordinaires, automorphes par le groupe cyclique, jusqu'à $n = 9$. a et $a - 1$ sont premiers avec n, b et c sont quelconques.

Pour toute valeur de n on a d'abord la solution:

$$x \times y = a(x - y) + y + b$$

Pour $n = 7$, on a, en plus:

$$x \times y = \pm(x - y + c)^4 + 4(x - y + c) + y + b.$$

Pour $n = 9$, on trouve 28 solutions, aux isotopies D et E près, dont

$$x \times y = 3(x - y)^2 + 2x - y$$

et

$$x \times 0 = 1 + 6\binom{x}{1} + 7\binom{x}{2} + 2\binom{x}{3} + 5\binom{x}{4} + 3\binom{x}{6} + 6\binom{x}{7}.$$

Quand n est pair, il faut avoir recours à des fonctions irrationnelles. Exemple, $n = 4$:

$$f(x) = x \times 0 = \sqrt{(17x^2 - 41x + 24)/6}$$

représente le quasigroupe du N° 8.

III. Composition des quasigroupes

15. *Si $f(x)$ est un quasigroupe d'ordre p automorphe par le groupe cyclique et si $\rho_x(y)$, $x = 0, 1, \ldots, p-1$, sont p quasigroupes d'ordre m encore automorphes par le groupe cyclique, alors:*

$$F(x + py) = f(x) + p\rho_x(y) \pmod{n}$$

est un quasigroupe d'ordre $n = mp$, automorphe par le groupe cyclique.

Exemple: $f = (01)\,(2)$; $\rho_0 = (0214)\,(3)$; $\rho_1 = (02)\,(34)\,(1)$; $\rho_2 = (0412)\,(3)$.

$X = x + 3y =$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$F(X) =$	7	6	14	13	3	8	4	0	2	10	12	11	1	9	5

Quand x décrit le champ $(0, p-1)$ et y le champ $(0, m-1)$, la fonction $f(x)$ d'une part et chacune des fonctions $\rho_x(y)$ de l'autre décrivent respectivement les mêmes champs. Donc F prend toutes les valeurs de 0 à mp − 1 et il en est de même de $x - F(x)$.

16. *Dans les conditions du* N° 15, *il existe une partition régulière sur le quasigroupe $F(X)$. Les p cosets sont les ensembles $x + py$, où y décrit les m valeurs $0,1, \ldots, m-1$. A chaque valeur de x correspond un coset; le système des représentants est $x = 0,1, \ldots, p-1$. Le quasigroupe quotient est isomorphe au quasigroupe $f(x)$. Pour que la partition définisse un diviseur normal* **(4)** *il faut et il suffit que $f(x)$ soit idempotent; tous les cosets sont alors diviseurs normaux.*

Il est d'abord évident que les ensembles $x + py$ sont disjoints car si

$$a + py \equiv b + py' \pmod{mp},$$

on aura

$$a - b \equiv 0 \pmod{p},$$

donc

$$a = b,$$

puisque a et b sont plus petits que p.

Soient $a + py$ un élément du coset $x = a$ et $b + py'$ un élément du coset $x = b$. En notant $(\times)$, $(*)$ et $(\cdot)$ les opérations de f, ρ_{a-b} et $F = f + p\rho$, leur produit sera:

$$P = (a + py) \cdot (b + py') = [a - b + p(y - y')]\cdot 0 + b + py'$$
$$= F[a - b + p(y - y')] + b + py'$$
$$= f(a - b) + p\rho_{a-b}(y - y') + b + py'$$
$$= (a - b) \times 0 + b + p[(y - y') * 0 + y']$$
$$P = a \times b + p(y * y') = c + py''.$$

Ainsi P appartient au coset représenté par $x = a \times b = c$.

La loi de composition des cosets est la même que celle de leurs représentants et il existe un homomorphisme canonique de F sur f.

Si f n'a aucun élément idempotent, aucun coset n'est fermé. Si un coset est diviseur normal, tous le seront, car si $a \times a = a$, tous les éléments de $f(x)$ seront idempotents, à cause de l'automorphisme $(a + 1) \times (a + 1) = a + 1$.

Exemple. Le quasigroupe examiné au N° 15 et représenté, avec la notation du N° 9, par:

$$F(x) = (0,7)\ (1, 6, 4, 3, 13, 9, 10, 12)\ (2, 14, 5, 8)\ (11)$$

a trois cosets

$$0, 3, 6, 9, 12;\quad 1, 4, 7, 10, 13;\quad \text{et}\quad 2, 5, 8, 11,14,$$

mais $f(x)$ n'ayant aucun élément idempotent, aucun coset n'est diviseur normal; il y a seulement partition régulière et homomorphisme naturel du quasigroupe F sur le quasigroupe quotient, isomorphe à f.

Soit au contraire le quasigroupe défini par:

$$p = 3; \mathrm{m} = 5; f(x) = (12)\ (0); \rho_0 = \rho_1 = \rho_2 = (0)\ (1243),$$

d'où:

$$F = (1,2)\ (3, 6, 12, 9)\ (4, 8, 13, 11)\ (5, 7, 14, 10)\ (0)$$

(Notation du N° 9).

Comme $f(x)$ est idempotent, tous les cosets sont diviseurs normaux.

17. *Si $\rho_z(y)$ ne dépend plus de x, et si l'on considère, pour l'ensemble de tous les nombres entiers (p = 1,2,3, . . . ,) tous les quasigroupes $f(x)$ (mod. p) automorphes par le groupe cyclique d'ordre p, les fonctions f forment un semi-groupe* **(5)** *non commutatif, avec unité bilatère, par rapport à la loi de composition*

$$f*\rho = f + p\rho.$$

Cet ensemble est fermé d'après le N° 15. Il est associatif car, si f, ρ et ψ sont trois fonctions, mod. p, q et r respectivement, on a

$$(f*\rho)*\psi = (f + p\rho)*\psi = f + p\rho + pq\psi \qquad (\text{mod. } pqr),$$
$$f*(\rho*\psi) = f*(\rho + q\psi) = f + p\rho + pq\,\psi \qquad (\text{mod. } pqr).$$

L'unité est la fonction $f = 0$, correspondant au quasigroupe du 1er ordre $0*0 = 0 = u$; elle est bilatère car $f*u = u*f = f$.

18. *Si $p,p',\ldots$ sont des entiers positifs quelconques et si $M = \{p,p',\ldots\}$ est le semi-groupe multiplicatif engendré par ces entiers, l'ensemble des fonctions $f(x)$* (mod. q), $q \in M$, *est un semi-groupe, diviseur du semi-groupe défini au N° précédent.*

Car il est fermé et contient l'unité; il est associatif. En particulier l'ensemble, au sens de la multiplication, $*$, définie au N° 17, des puissances d'une fonction unique $f(x)$, d'ordre q, est isomorphe au semi-groupe additif des entiers. Si l'on pose:

$$f^{*p} = 0 + f + qf + q^2f + \ldots + q^{p-1}f,$$

alors,

$$f^{*p} * f^{*r} = f^{*(p+r)}.$$

IV. **Recherches sur l'automorphe (8, p. 40) de f(x)**

19. *Quand un quasigroupe $Q(\times)$, fini ou non, automorphe par toutes les transformations C; $x \to x + h$, est engendré par un de ses éléments, son automorphe se réduit au seul groupe cyclique C.*

Soit $Q = \{a\}$; à cause de l'automorphisme $x \to x - a$, Q est aussi engendré par l'élément zéro. Soit

$$\binom{x}{x'} = H$$

un automorphisme quelconque de Q, et v l'image de 0 par H. On montre par induction que, pour tout x dans Q, on aura: $xH = x' = x + v$.

Soient a et b deux élément de Q pour lesquels on suppose:

$$a' = a + v \text{ et } b' = b + v.$$

En vertu de l'automorphisme C:

$$a' \times b' = (a' - b') \times 0 + b' = (a - b) \times 0 + b + v,$$

et comme

$$a \times b = (a - b) \times 0 + b,$$
$$a' \times b' = a \times b + v.$$

La propriété étant vraie pour $a = b = 0$, et Q étant engendré par 0, la propriété est générale et $H \in C$.

20. En particulier, *si l'une des trois équations*: $0 \times 0 = x$; $x \times 0 = 0$; $0 \times x = 0$ *a pour solution un nombre premier avec n, le quasigroupe $Q(\times)$, d'ordre n, automorphe par le groupe cyclique C, admet C pour automorphe.*

21. *Si un quasigroupe $Q(\times)$ est automorphe par le groupe cyclique, si $0 \times 0 = a$ et si $x \to x'$ est un automorphisme quelconque de Q, on a, pour toute valeur de k;*

$$(ka + r)' = ka + r'.$$

La proposition est évidente si k est nul et on la généralise sans peine par induction.

22. *Si $Q(\times)$ est un quasigroupe d'ordre n, automorphe par le groupe cyclique $C = \{c\}$, $c = 0, 1, \ldots, n-1$, tout diviseur propre D de Q est composé des éléments de une ou plusieurs suites;*

$$i, i + p, i + 2p, \ldots, i + kp, \ldots, i + n - p,$$

où $p|n$ et $i < p$; la valeur de p restant la même pour toutes les suites de D. Il existe p telles suites, disjointes, correspondant aux p valeurs de $i (i = 0,1, \ldots, p-1)$. Enfin, D est automorphe par le diviseur $\{c^p\}$ du groupe cyclique C.

Puisque C opère transitivement sur Q, tout élément de C ne peut pas être un automorphisme de D. Comme les conditions $c^a D = D$ $c^b D = D$ impliquent que $c^{a-b} D = D$, le sous-ensemble des éléments de C, qui sont des automorphismes de D, forme un sous-groupe. C étant cyclique, ce sous-groupe est de la forme $\{c^p\}$, où $p|n$ et est le plus petit entier satisfaisant à $c^p D = D$

Si D contient l'élément i, il contiendra les $n/p = m$ éléments distincts:

$$i, i + p, i + 2p, \ldots, i + kp, \ldots, i + n - p. \tag{4}$$

L'automorphisme $x \to x + r$, $(r < p)$ transformera D en un diviseur D', pouvant coincïder avec D, qui contiendra les éléments:

$$i + r, p + i + r, \ldots, kp + i + r, \ldots, n - p + i + r. \tag{5}$$

On voit immédiatement que les suites (5) correspondant à des valeurs distinctes de r sont disjointes. Elles réalisent une partition des n éléments de Q, épuisant l'ensemble de ces éléments, puisqu'elles sont disjointes, au nombre de p, et composées de n/p éléments chacune.

23. *Si un quasigroupe $Q(\times)$ d'ordre n, automorphe par le groupe cyclique, $x \to x + h$, n'est engendré par aucun de ses éléments.*

(1) *Chaque élément de Q engendre un diviseur d'ordre constant m $(mp = n)$.*

(2) *Le nombre des diviseurs distincts ainsi définis est p et chacun d'eux est engendré par n'importe quel de ses éléments.*

(3) *Ces p diviseurs sont disjoints, isomorphes entre eux et l'on passe de l'un à tous les autres par les transformations:*

$$x \to x + h \ (h = 0,1,2, \ldots, p-1).$$

(4) *Le diviseur engendré par l'élément* 0 *est*: $D = 0, p, 2p, 3p, \ldots, n - p$, *et on l'obtient en multipliant par* p, *sans changer la loi de composition, un quasigroupe quelconque d'ordre* m, *automorphe par le seul groupe cyclique d'ordre* m, *c'est-à-dire engendré par un seul de ses éléments.*

(5) *Tous ces diviseurs ont le même automorphe, à savoir le groupe cyclique d'ordre* m:

$$x \to x + kp,\ k = 0,1,2, \ldots, m - 1.$$

Preuve. Soit un quasigroupe $Q(\times)$ d'ordre n, non monogène, et automorphe par le groupe cyclique C: $x \to x + h$; $h = 0,1,2, \ldots, n - 1$. Soit D le sous-quasigroupe propre engendré par l'élément zéro: $\{0\} = D \subset Q$. Si l'on fait subir à D toutes les transformations C, à chaque valeur i de h correspondra un diviseur D_i, isomorphe à D:

$$D_i = D\begin{pmatrix} x \\ x + i \end{pmatrix},$$

et D_i sera engendré par l'élément i de la même manière que D l'était par l'élément 0. On obtiendra ainsi n diviseurs.

Soit a un élément non nul de D, $D_a = \{a\}$. Comme $a \in D$, $D_a \subset D$. Mais $D_a \cong D$, donc $D_a = D$. Ainsi chacun des n diviseurs D_i est engendré par n'importe quel de ses éléments.

Soit p le plus petit élément non nul de $D = \{0\}$. Donc

$$D_p = \{p\} = D = D\begin{pmatrix} x \\ x + p \end{pmatrix}$$

Par suite D contient, outre l'élément zéro, les éléments $p,2p,3p, \ldots, kp; \ldots$ (mod. n), dont le plus petit (mod. n) est p par hypothèse. On voit facilement que $p|n$ et $kp < n$; on peut poser $mp = n$; $k < m$. Supposons alors que D, en plus de $0,p,2p, \ldots, (m - 1)p$, contienne un autre élément q, non multiple de p, par exemple:

$$kp < q < kp + p \leqslant n,$$

l'égalité correspondant au cas où $k = m - 1$.

S'il en était ainsi, le diviseur:

$$D' = D\begin{pmatrix} x \\ x - kp \end{pmatrix}$$

contiendrait l'élément zéro et coïnciderait avec D. D'autre part, il contiendrait l'élément $q - kp$. Or

$$0 < q - kp < p$$

et, contrairement à l'hypothèse, p ne serait pas le plus petit élément non nul de D. Ainsi, $D = 0, p, 2p, \ldots, kp, \ldots, (m - 1)p$.

Chaque diviseur D_i est d'ordre m et le nombre des D_i distincts est p.

D'après le N° précédent ces diviseurs sont disjoints. Si l'on remplace dans D tous les éléments $0,p,2p,\ldots,n-p$ par leurs quotients par p:

$$0,1,2,\ldots,m-1,$$

en respectant la loi de composition de Q, on obtient évidenment un quasi-groupe isomorphe à D par $x \to (x/p)$, dont l'automorphe se réduit (N° 19) au groupe cyclique d'ordre m: $x \to x+k$, $(k = 0, 1,2,\ldots,m-1)$ puisqu'il est engendré par l'un quelconque de ses éléments.

24. *Si un quasigroupe $Q(\times)$ est automorphe par le groupe cyclique et si $D_0, D_1, D_2, \ldots, D_{p-1}$ sont les diviseurs d'ordre $m = n/p$ engendrés par les divers éléments de Q, tout automorphisme T de Q projette chaque diviseur D_i sur un diviseur D_j, l'isomorphisme induit par T entre D_i et D_j étant:*

$$x \to x + j - i \quad (iT = j)$$

L'automorphe de Q induit sur l'ensemble des indices $0,1,2\ldots,p-1$ un groupe de transformations.

Preuve. Soit T un automorphisme de Q et soit a l'image de zéro.

$$0T = a.$$

Alors tous les produits engendrés par 0 se projettent sur les produits engendrés de la même manière par a et tous les éléments de D_0 ont des images bien déterminées et définies par la seule donnée de l'image de 0.

Si $0 \times 0 = kp$, d'après l'automorphisme,

$$a \times a = kp + a,$$

de sorte que,

$$D_0T = D_a.$$

Plus généralement si i a pour image j, l'image de $\{i\}$ sera évidenment $\{j\}$, car si

$$iT = j = i + h,$$

d'après le N° 21

$$(kp + i)T = kp + iT = kp + i + h, \quad D_iT = D_{i+h} = D_i.$$

Tout automorphisme T de Q induit donc sur l'ensemble des diviseurs $D_0, D_1, D_2, \ldots, D_{p-1}$ une transformation t des indices $0,1,2,3,\ldots,p-1$. L'automorphe de Q induit sur l'ensemble $0,1,2,\ldots,p-1$ un groupe de substitutions.

25. *Dans les conditions du N° 24, si g est l'ordre du système minimal de générateurs de Q, l'ordre de l'automorphe de Q est un diviseur de $m^g p!/(p-g)!$.*

Si l'on choisit un élément quelconque dans chaque diviseur: $a_0 \in D_0$, $a_1 \in D_1, \ldots, a_i \in D_i, \ldots$ d'après le N° 24, tout automorphisme T de Q sera entièrement défini si l'on connaît les images:

$$a'_0 = a_0 T, \ldots, a'_i = a_i T, \ldots .$$

Mais ces images ne sont pas indépendantes enaénéral.

Soit $a_i, a_j, \ldots, a_k$ un système minimal de générateurs indépendants de Q (c'est-à-dire comprenant le plus petit nombre possible de générateurs tels que, par la suppression d'un de ces générateurs, les éléments restants engendrent seulement une partie propre de Q). Alors si les images des générateurs sont connues, il est clair que l'image d'un élément quelconque de Q sera définie. Le nombre des éléments de Q dont on peut se donner arbitrairement l'image est donc égal au nombre minimum des générateurs; soit g ce nombre.

On observera encore que ces g générateurs appartiennent nécessairement à g diviseurs distincts parmi les diviseurs $D_0, D_1, \ldots, D_{p-1}$ Car si deux d'entre eux appartenaient au même diviseur, D, on aurait évidemment un système de générateurs plus court en supprimant l'un d'eux.

Il y a
$$m^g \binom{p}{g} g! = m^g \, p!/(p-g)! = q$$
manières de choisir les images des g générateurs. Les substitutions ainsi obtenues forment un groupe d'ordre q, dont l'automorphe de Q est un diviseur, toutes ces substitutions n'étant pas nécessairement des automorphismes de Q.

Si $p = 1$, alors $g = 1$, $m = n = q$ et l'automorphe de Q est d'ordre n, conformément à la conclusion du N° 19.

26. *Le diviseur D_0 n'est pas nécessairement normal.*

Exemple.

Le quasigroupe:

$$S: F(x) = (0, 10)\,(1, 3, 11, 14, 8, 9, 7, 13, 12, 4)\,(2, 6)\,(5)$$

automorphe par le groupe cyclique du 15° ordre, a pour diviseurs

$$D_i = i, 5 + i, 10 + i \quad (i = 0, 1, 2, 3, 4).$$

Mais aucun de ces diviseurs n'est normal.

27. *Si Q est un quasigroupe d'ordre $n = mp$, automorphe par le groupe cyclique C_n, si $D_0, D_1, \ldots, D_{p-1}$ sont les p diviseurs d'ordre m engendrés par les divers éléments de Q, pour que D_0 (et par conséquent tout D_i) soit normal dans Q il faut et il suffit que Q soit défini par une fonction $F(i + py) = f(i) + p\,\rho_{ij}$ où $j = f(i)$ est un quasigroupe idempotent d'ordre p, automorphe par le groupe*

cyclique C_p et où $\rho_i(y)$, $(y = 0,1,2 \ldots, m-1)$ définit pour chaque valeur de i $(i = 0, 1, \ldots, p-1)$ un quasigroupe d'ordre m, automorphe par le groupe cyclique C_m.

On a alors $D_i \times D_0 = D_j$ et le quasigroupe quotient Q/D est isomorphe à $f(i)$ par $D_i \to i$.

Preuve. Si D est normal, tout D_i est normal et il est évident que, pour toute valeur de i, les produits de tous les éléments de D_i par 0 doivent appartenir à un seul diviseur, par exemple D_j. Réciproquement, si cela a lieu, il est facile de montrer que tous les D_i sont normaux. En effet, par définition (23), on a:

$$D_i = D_0\begin{pmatrix} x \\ x+i \end{pmatrix}.$$

Supposons: $D_i \times 0 = D_j$, alors, à cause de l'automorphisme

$$D_{i+kp} \times kp = D_{j+kp}.$$

Mais $D_{i+kp} = D_i$, puisque

$$D_{i+kp} = D_i\begin{pmatrix} x \\ x+kp \end{pmatrix}$$

et que $D_i = i, i+p, \ldots,$ donc

$$D_i \times kp = D_j,$$

quelque soit k, ce qui signifie

$$D_i \times D_0 = D_j,$$

et par suite

$$D_{i+h} \times D_h = D_{j+h}.$$

Donc D_0 sera bien normal si, pour toute valeur de i $(i = 0, 1,2, \ldots, p-1)$ on a

$$D_i \times 0 = D_j.$$

Cette relation établit une correspondance biunivoque entre chaque élément de D_i: $i + kp \in D_i$ et un élément: $j + z_k p$ de D_j. Reste à exprimer que cette transformation satisfait pour tout $x \in Q$, à la condition:

$$x - x \times 0 = y - y \times 0 \rightleftarrows x = y \qquad \text{(mod. } n\text{)}$$

(On observera que D est toujours normal si $m = n/p = 3$.)

En s'occupant d'abord des $x \in D_i$, on aura:

$$\begin{array}{llllll} x = i, & i+p, & i+2p, \ldots, & i+kp, \ldots, & i+mp-p. \\ F = j+z_0p, & j+z_1p, & j+z_2p, \ldots, & j+z_kp, \ldots, & j+z_{m-1}p. \end{array}$$

$$x - F = i - j - z_0p, i - j - (z_1 - 1)p, i - j - (z_2 - 2)p, \ldots,$$
$$i - j - (z_k - k)p, \ldots, i - j - (z_{m-1} - m + 1)p.$$

Pour que ces différences soient toutes distinctes deux à deux, il faut que, en posant:

$$y = 0, 1, 2, 3, \ldots, k, \ldots, m-1.$$

et

$$\rho_i(y) = z_0, z_1, z_2, z_3 \ldots, z_k \ldots, z_{m-1},$$

la fonction $\rho_i(y)$ satisfasse à la condition:

$$y - \rho_i(y) = y' - \rho_i(y') \rightleftarrows y = y' \qquad \text{(mod. } m\text{)}$$

Cela exprime que ρ définit un quasigroupe d'ordre m, automorphe par le groupe cyclique C_m et d'ailleurs quelconque.

Si toutes les différences $i - j + (k - z_k)p$ sont distinctes elles seront, à l'ordre près, égales à:

$$(6) \qquad i - j, i - j + p, i - j + 2p, \ldots, i - j + (m-1)p.$$

En considérant maintenant toutes les valeurs de x, il faut exprimer que la suite (6) n'a aucun terme commun avec chacune des suites analogues:

$$(7) \qquad i' - j', i' - j' + p, i' - j' + 2p, \ldots, i' - j' + (m-1)p,$$

où $i, j, i', j' < p$.

Or il est facile de voir que, pour que (6) et (7) n'aient aucun terme commun il faut et il suffit que:

$$i - j \not\equiv i' - j',$$

car: [1], si $i - j = i' - j'$, alors les suites (6) et (7) ont tous leurs termes deux à deux égaux et par conséquent ne sont pas distinctes; [2], si (6) et (7) ont un terme commun, par exemple:

$$i - j + kp \equiv i' - j' + k'p \qquad \text{(mod. } n\text{)},$$

on en déduira

$$(i - j) - (i' - j') \equiv (k' - k)p \qquad \text{(mod. } mp\text{)}$$

ce qui exige

$$i - j \equiv i' - j' \qquad \text{(mod. } p\text{)}$$

et comme i et j sont plus petits que p,

$$i - j = i' - j'.$$

Finalement il faut que la transformation $i \to j$ satisfasse à

$$i - j \not\equiv i' - j', \qquad i \not\equiv i',$$

ou, en posant $j = f(i)$,

$$i - f(i) \not\equiv i' - f(i'), \qquad i \not\equiv i'.$$

Donc la fonction $f(i)$ définit encore un quasigroupe, d'ordre p, automorphe par le groupe cyclique C_p. D'ailleurs ce quasigroupe est idempotent car $D_i \times D_i = D_i$.

La réciproque résulte du N° 15 et la dernière partie du N° 16.

28. *Dans les conditions du N° 27, si P est l'ordre de l'automorphe de $f(i)$, le nombre des automorphismes de Q est un diviseur de Pm^g où g désigne comme au N° 25 le nombre minimum de générateurs indépendants de Q.*

Tout automorphisme de Q est défini par un ensemble de p substitutions

$$s_i : x \to x + kp + j - i \qquad (k = 0, 1, 2, \ldots, m-1)$$

où $i \to j$ est un automorphisme de $f(i)$ et où la substitution s_i opère sur un seul diviseur D_i.

D'autre part, d'après le N° 24, tout automorphisme T de Q projette un D_i sur un D_j, c'est-à-dire permute les D_i entre eux. D étant normal, T induit sur le quasigroupe quotient $f(i)$ d'ordre p un automorphisme t. Et, si $iT = j$, l'automorphisme T induit un isomorphisme de D_i sur D_j.

Le nombre des automorphismes t est au plus égal à celui des automorphismes du quasigroupe $f(i)$, soit P.

Comme tous les D sont monogènes, chacun a pour automorphe le seul groupe cyclique d'ordre m.

$$x \to x + kp \qquad (k = 0, 1, \ldots, m-1).$$

Si R est un isomorphisme de D_i avec D_j,

$$D_i R = D_j,$$

comme on a:

$$D_i \begin{pmatrix} x \\ x + kp \end{pmatrix} = D_i$$

et par conséquent

$$D_i \begin{pmatrix} x \\ x + kp \end{pmatrix} R = D_j,$$

tous les autres automorphismes seront compris dans la formule:

$$\begin{pmatrix} x \\ x + kp \end{pmatrix} R.$$

Or D_j est isomorphe de D_i par $x + i \to x + j$, c'est-à-dire:

$$\begin{pmatrix} x \\ x + j - i \end{pmatrix}.$$

Donc tous les isomorphismes de D_i avec D_j sont exprimés par les transformations:

$$S = \begin{pmatrix} x \\ x + kp \end{pmatrix}\begin{pmatrix} x \\ x + j - i \end{pmatrix} = \begin{pmatrix} x \\ x + kp + j - i \end{pmatrix}$$

i, j fixes; $k = 0, 1, \ldots, m - 1$.

Pour chaque D_i il existe donc m substitutions, correspondant aux m valeurs de k et projetant D_i sur D_j.

L'automorphisme T de Q induit un isomorphisme de D_i avec D_j compris dans les m substitutions S ci-dessus. Cela peut être répété pour les p diviseurs D_i, mais si Q a g générateurs, le choix de k n'est libre que pour ces g générateurs et finalement le nombre total des automorphismes de Q est un diviseur de Pm^g, ordre du groupe défini par les t et par les S.

Exemple I.

$$F = (0, 3)\ (1, 5)\ (2, 4)\ (6)\ (7, 8)$$
$$p = m = 3;\ D_0 = 0, 3, 6;\ D_1 = 1, 4, 7;\ D_2 = 2, 5, 8$$
$$S_0 = \begin{pmatrix} x \\ x \end{pmatrix};\qquad S_1 = \begin{pmatrix} x \\ x + 3 \end{pmatrix};\qquad S_2 = \begin{pmatrix} x \\ x + 6 \end{pmatrix}$$
$$T = (0)\ (3)\ (6) \cdot (147) \cdot (285)$$

Exemple II.

$$f = (0)\ (1243);\ \rho_0 = \rho_2 = \rho_3 = (02)\ (1);\ \rho_1 = (01)\ (2)$$
$$F = (0, 10)\ (1, 7, 9, 13)\ (2, 14, 8, 6)\ (3, 11, 12, 4)\ (5) = f + 5\ \rho_i(y).$$

Ce quasigroupe est engendré par (0, 1). Tout automorphisme est défini par les images de 0 et de 1. L'automorphe de Q est donc d'ordre 9. 5!/3! = 180 ou un diviseur de 180.

En écrivant i pour D_i, le quasigroupe quotient devient $f(i)$ lui-même. Son automorphe est du 20ème ordre; il contient C_5 et le sousgroupe du 4ème ordre $\{(1243)\}$.

CITATIONS

1. A. A. Albert, *Non associative Algebras*, Ann. Math. (2), *43* (1942), 696.
2. B. A. Haussmann, O. Ore, *Theory of Quasigroups*. Amer. J. Math., *59* (1937), 983.
3. C. Jordan, *Traité des substitutions* (Paris, 1870).
4. F. Kiokemeister, *A theory of normality for quasi-groups*, Amer. J. Math., *70* (1948), 100–102.
5. D. Rees, *On semi-groups*, Proceed. Camb. Phil. Soc., *36* (1940), 387.
6. A. Sade, *Quasigroupes* (Marseille, 1950).
7. ———, *Contribution à la théorie des quasigroupes*, CR Acad. Sci. Paris, *237* (1953), 420–422.
8. G. Scorza, *Gruppi Astratti* (Roma, 1942).

Lycée Perier, Marseille .

SOME GENERALIZATIONS OF BURNSIDE'S THEOREM

N. A. WIEGMANN

1. Introduction. Burnside's Theorem in the theory of group representations states that a necessary and sufficient condition that a semigroup of matrices of degree n over the complex field be irreducible is that the semigroup contain n^2 linearly independent matrices. In the course of dealing with sets of matrices with coefficients in a division ring, Brauer **(1)** obtained a generalization of this theorem which concerned irreducible semigroups with elements in a division ring. In the present work irreducible semigroups of matrices with elements in the field of real quaternions are considered and generalizations of Burnside's Theorem of a more specific nature are obtained by using certain properties of matrices with such elements.

The following facts and terminology may be briefly noted. Let $\mathfrak{A}$ be a semigroup (relative to multiplication) of quaternion matrices of degree n. By the l-rank (i.e., left rank) of $\mathfrak{A}$ is meant the maximum number of left linearly independent matrices in $\mathfrak{A}$; r-rank has a corresponding meaning. If every matrix A of $\mathfrak{A}$ is of either form

$$\begin{bmatrix} A_1 & O \\ A_3 & A_4 \end{bmatrix} \text{ or } \begin{bmatrix} A_1 & A_2 \\ O & A_4 \end{bmatrix}$$

where A_1 is $m \times m$, $m < n$, the semigroup is said to be decomposed. If $\mathfrak{A}$ is such that there exists a non-singular quaternion matrix P such that the set $P\mathfrak{A}P^{-1}$ is decomposed, then $\mathfrak{A}$ is said to be reducible; if not $\mathfrak{A}$ is said to be irreducible. According to a result **(1, 4.4B)** of Brauer's, Schur's Lemma holds for semigroups of matrices with elements in a division ring: If $\mathfrak{A}$ and $\mathfrak{B}$ are irreducible semigroups which are intertwined by a matrix P, then P is either 0 or non-singular. This result could be obtained in the real quaternion case by paralleling Schur's proof while using the result **(3,** Theorem 10**)** that for every quaternion matrix P there exist unitary quaternion matrices U and V such that UPV is a real diagonal matrix with non-negative diagonal elements. It can also be shown that the following is true: If two semigroups $\mathfrak{A}$ and $\mathfrak{B}$ of quaternion matrices are intertwined by a matrix P and if $\mathfrak{B}$ is irreducible, then either $P = 0$ or $\mathfrak{A}$ contains $\mathfrak{B}$ as an irreducible component.

2. The form of a matrix commutative with an irreducible representation.

THEOREM 1. *If $\mathfrak{A}$ is an irreducible semigroup of quaternion matrices, then any matrix M which commutes with every element of $\mathfrak{A}$ is of the form $M = P^{-1}(kI)P$ where k is a complex number and P is a non-singular quaternion matrix.*

Received May 1, 1956.

Let M be a quaternion matrix with the given property. It is known **(3**, Theorem 1**)** that there exists a non-singular matrix P such that $PMP^{-1} = J$ is in Jordan normal form with complex elements along the diagonal.

If M is non-derogatory **(3)**, then each A in $\mathfrak{A}$ is such that $PAP^{-1} = B_1 \dotplus B_2 \dotplus \ldots \dotplus B_m$ where each B_t is triangular **(3**, pp. 195**)**, but this contradicts the assumption on $\mathfrak{A}$.

If M is derogatory, J is of the form kI, where k is a complex number, or it is not. If not, $J = J_1 \dotplus J_2 \dotplus \ldots \dotplus J_t$ where this form may be assumed to be such that J_t contains only and all diagonal elements λ_t which are characteristic roots of M, where $\lambda_i \neq \lambda_j$ for $i \neq j$ and where $\lambda_i \neq \bar{\lambda}_j$, for $i \neq j$, the latter according to the Jordan form designated for M in **(3)**. Now either J has one characteristic root or it has more than one; i.e., either $t = 1$, above, or $t > 1$ in J_t. These cases are now considered.

Let $J = J_1 \dotplus J_2 \dotplus \ldots \dotplus J_t$, $t > 1$. Let A be any element of $\mathfrak{A}$ and let $PAP^{-1} = X = (x_{ij})$. It will be shown that

$$X = \begin{bmatrix} X_1 & X_2 \\ 0 & X_3 \end{bmatrix},$$

where X_1 has the same order as that of J_1. Let j be the order of J_1 and let the first block (in the direct sum which J_2 represents) of the form

$$\begin{bmatrix} \lambda_2 & 1 & 0 & & \\ & \lambda_2 & & & \\ & & \ldots & & 1 \\ 0 & & & & \lambda_2 \end{bmatrix}$$

be of order k. Then we have a series of relations of the form

(i) $\quad \lambda_2 x_{sp} + x_{s+1,p} = x_{sp}\lambda_1, \quad \lambda_2 x_{j+k,p} = x_{j+k,p}\lambda_1,$

or

(ii) $\quad \lambda_2 x_{sp} + x_{s+1,p} = x_{s,p-1} + x_{sp}\lambda_1, \quad \lambda_2 x_{j+k,p} = x_{j+k,p-1} + x_{j+k,p}\lambda_1,$

where, for a fixed p chosen from $p = 1, 2, \ldots, j$, either (i) or (ii) holds (but not both); where for $p = 1$, (i) holds; and where if $k > 1$, $j < s < j + k$, and where if $k = 1$, only the second relations in (i) and (ii) hold. Now one of two cases occurs:

(a) At least one λ_t of M is real and there is no loss in generality in assuming λ_1 is such a real root. It follows from the second relation in (i) that $x_{j+k,1} = 0$, and from the first that $x_{s1} = 0$ for $j < s < j + k$. For any other p for which (i) holds, the same is true so that $x_{sp} = 0$ for $j < s \leqslant j + k$ for these p. Consider the first p for which (ii) holds; then in (ii), $x_{s,p-1} = 0 = x_{j+k,p-1}$ and so $x_{j+k,p} = 0$ from the second relation, and from the first $\lambda_2 x_{j+k-1,p} + 0 = 0 + x_{j+k-1,p}\lambda_1$ so that $x_{j+k-1,p} = 0$ and, in turn, $x_{sp} = 0$ for $j < s \leqslant j + k$ for this p. If the next p for which (ii) holds is treated in a similar fashion it follows that $x_{sp} = 0$ for $j < s \leqslant j + k$, and $1 \leqslant p \leqslant j$.

(b) If all λ_t are non-real complex, it follows from the second relation in (i) that $x_{j+k,1} = 0$ (since $\lambda_2 \neq \lambda_1$ and $\bar{\lambda}_2 \neq \lambda_1$) and from the first relation that $x_{s1} = 0$ for $j < s < j + k$. For any p for which (i) is true, the same result holds. Considering the first p for which (ii) holds, it follows as in (a) that all $x_{sp} = 0$ for this p and the same range of s so that $x_{sp} = 0$ here, also, for $j < s \leqslant j + k$ and $1 \leqslant p \leqslant j$. Now for any subsequent such blocks in J_2 or for any such blocks in $J_3, \ldots, J_t$ a similar procedure applies so that $x_{lp} = 0$ for $1 \leqslant p \leqslant j$ and for $j < l \leqslant n$ (where X is $n \times n$). So X has the above form which contradicts the assumption of the theorem.

Let $J = J_1$ so that only λ_1 appears along the diagonal where there must be an element 1 to the right of at least one λ_1. Let $J_1 = J_{11} \dotplus J_{22} \dotplus \ldots \dotplus J_{kk}$ where each J_{tt} is either the single element λ_1 or a matrix with λ_1 along the diagonal and 1 above each such λ_1 after the first, and where J_{11} is definitely of the latter form. Let X_{tt} have the same order as J_{tt} and let

$$X = \begin{bmatrix} X_{11} & X_{12} \ldots X_{1k} \\ & \ldots \\ X_{k1} & X_{k2} \ldots X_{kk} \end{bmatrix},$$

so that, from the assumptions, X_{11}, X_{12}, X_{21}, and X_{22} at least appear and are not vacuous and $k < n$.

Now it will be shown that it follows from $JX = XJ$ that each X_{tj} is a triangular matrix with zeros below the main diagonal; it is to be noted here that when some X_{tj}, $i \neq j$, is not square, the latter statement will be taken to mean that any element $x_{tj} = 0$ when $i > j$. Then it follows that $J_{tt}X_{tj} = X_{tj}J_{jj}$ $(i, j = 1, 2, \ldots, k)$. When $i = j$, since J_{tt} is non-derogatory, X_{tt} is in triangular form $(i = 1, 2, \ldots, k)$. When $i \neq j$, consider for example X_{12} as a typical case: $J_{11}X_{12} = X_{12}J_{22}$, and let these be of order $s \times s$, $s \times t$, and $t \times t$ respectively. Equating the elements in the first column (and dropping the subscript on λ_1):

(i) $$\lambda x_{t1} + x_{t+1,1} = x_{t1}\lambda\,, \quad 1 \leqslant i < s\,, \quad \lambda x_{s1} = x_{s1}\lambda\,;$$

and from the other columns there result:

(ii) $$\lambda x_{lj} + x_{l+1,j} = x_{l,j-1} + x_{lj}\lambda\,, \quad \lambda x_{sj} = x_{s,j-1} + x_{sj}\lambda\,,$$
$$1 \leqslant l < s \quad \text{and} \quad 1 < j \leqq t\,,$$

where the following is to be noted: if X_{tj} consists of one column, only (i) applies, and if X_{tj} consists of only one row, the second relations only in (i) and (ii) apply. (In the latter case X_{tj} is already in triangular form.) If λ is real, from the first relation in (i), $x_{t+1,1} = 0$ for $1 \leqslant i < s$; from the first relation in (ii) $x_{l+1,j} = x_{l,j-1}$ for $1 \leqslant l < s$, $1 < j \leqq t$. Together these show $x_{tj} = 0$ for $i > j$. If λ is non-real complex, from the second relation in (i) x_{s1} must be complex and from the first $x_{s1} = 0$ and also, in turn,

$$x_{s-1,1} = x_{s-2,1} = \ldots = x_{21} = 0.$$

Since $x_{s1} = 0$, from the second relation in (ii) x_{s2} is complex and from the first $x_{s2} = 0$; similarly $x_{s-1,2} = \ldots = x_{22} = 0$. Continuing with all elements x_{sp} in the last row of X_{12}, it will be seen that $x_{ij} = 0$ for $i > j$.

For each A in $\mathfrak{A}$, $PAP^{-1} = X$ is of the above type, sectioned as above, and having each X_{ij} in triangular form as described. Form the matrices

$$\begin{bmatrix} X_{11} \\ \cdot \\ \cdot \\ \cdot \\ X_{k1} \end{bmatrix}, \begin{bmatrix} X_{12} \\ \cdot \\ \cdot \\ \cdot \\ X_{k2} \end{bmatrix}, \ldots, \begin{bmatrix} X_{1k} \\ \cdot \\ \cdot \\ \cdot \\ X_{kk} \end{bmatrix}.$$

Let X_{ij} have row order r_{ij} and column order c_{ij}; then the first columns of each of the above matrices have zeros in all the same positions and have, possibly, non-zero elements only in the 1st, the $(r_{11} + 1)$st, the $(r_{11} + r_{21} + 1)$st, . . . , the $(r_{11} + r_{21} + \ldots + r_{k-1,1} + 1)$st positions. By column operations on the right of X, the following columns of X may be interchanged:

the 2nd and the $(c_{11} + 1)$st,
the 3rd and the $(c_{11} + c_{12} + 1)$st,
the 4th and the $(c_{11} + c_{12} + c_{13} + 1)$st,
. .
. .
. .
the kth and the $(c_{11} + c_{12} + \ldots + c_{1,k-1} + 1)$st.

(Note that $c_{11} > 1$.) The resultant matrix is such that the first k columns have all-zero rows except for the 1st, the $(r_{11} + 1)$st, . . . , the $(r_{11} + r_{21} + \ldots + r_{k-1,1} + 1)$st. Now

$$r_{11} = c_{11},\ r_{21} = c_{12},\ r_{31} = c_{13},\ \ldots,\ r_{k-1,1} = c_{1,k-1}$$

so that if the same operations as above are now performed on the left on the rows, a similarity transformation will have been applied to X with the result that a matrix is obtained such that the first k columns have all-zero rows beyond the kth row; therefore, X is similar to a matrix

$$\begin{bmatrix} X_1 & X_2 \\ 0 & X_3 \end{bmatrix},$$

where X_1 is $k \times k$, $k < n$.

It follows then that $M = P^{-1}(kI)P$ where k is complex and P is quaternion.

3. Generalizations of Burnside's Theorem. If $\mathfrak{A}$ is a semigroup consisting of square matrices of degree n with quaternion elements, there may exist a non-singular (quaternion) matrix P such that $P\mathfrak{A}P^{-1} = \mathfrak{A}_1$ contains only matrices with complex elements, or there may not exist such a P. An example of the former may be obtained by forming $P^{-1}\mathfrak{A}_1P$ where $\mathfrak{A}_1$ is any

complex semigroup and P is a non-singular quaternion matrix. An example of the latter (other than the set of all $n \times n$ quaternion matrices) is the set, $\mathfrak{A}$, of all $n \times n$ unitary matrices with quaternion elements. This set is closed under matrix multiplication. There exists no P such that $P\mathfrak{A}P^{-1} = \mathfrak{C}$ is a complex set; this can be seen as follows: P cannot be complex since then $\mathfrak{A}$ would be complex. Also, in the notation of (3, p. 191) no $P = P_1 + jP_2$, P_1 and P_2 complex, $P_2 \neq 0$, can provide this. For if so, let $\mathfrak{U}$ be the set of all $n \times n$ complex unitary matrices; this is a semigroup of complex matrices which is irreducible under all complex similarity transformations. Since $\mathfrak{U}$ is a subset of $\mathfrak{A}$, P must be such that $P\mathfrak{U} = \mathfrak{C}_1P$ where $\mathfrak{C}_1$ is composed solely of complex matrices. From $(P_1 + jP_2)\mathfrak{U} = \mathfrak{C}_1(P_1 + jP_2)$ it follows that $P_1\mathfrak{U} = \mathfrak{C}_1P_1$ and $jP_2\mathfrak{U} = j\bar{\mathfrak{C}}_1P_2$ (where $\bar{\mathfrak{C}}_1$ consists of the complex conjugate of each matrix in $\mathfrak{C}_1$). From the first relation and from results in complex theory, either $P_1 = 0$ or $\mathfrak{C}_1$ contains $\mathfrak{U}$ as an irreducible component and since both $\mathfrak{C}_1$ and $\mathfrak{U}$ are $n \times n$ in dimension, $\mathfrak{C}_1$ is irreducible also, relative to the complex field. By Schur's Lemma in complex theory either $P_1 = 0$ or P_1 is non-singular. The latter must hold for if $P_1 = 0$, $P_2\mathfrak{U} = \bar{\mathfrak{C}}_1P_2$ and since $\mathfrak{U}$ and $\bar{\mathfrak{C}}_1$ are irreducible, either $P_2 = 0$ also (not possible) or P_2 is non-singular in which case $\mathfrak{A} = P^{-1}\mathfrak{C}P = P_2^{-1}j^{-1}\mathfrak{C}jP_2 = P_2^{-1}\bar{\mathfrak{C}}P_2$ is complex. Therefore P_1 is non-singular. From $jP_2\mathfrak{U} = j\bar{\mathfrak{C}}_1P_2$, either $P_2 = 0$ (not possible) or P_2 is non-singular. Then $P = P_1 + jP_2$ where both P_1 and P_2 are non-singular. But this is not possible because, for example, since $(ij)I$ and $2^{-\frac{1}{2}}(i + j)I$ are elements of $\mathfrak{A}$, they must be similar under the P^{-1}, P transformation to complex matrices C_1 and C_2, respectively. From the first relation it follows that

$$\bar{P}_2P_1^{-1} = P_1\bar{P}_2^{-1},$$

and from this and the second, the contradictory fact that $C_2 = 0$ would result.

THEOREM 2. *Let $\mathfrak{A}$ be an irreducible semigroup of quaternion matrices. Then $M = P(kI)P^{-1}$, where k is non-real complex, commutes with each element of $\mathfrak{A}$ if and only if $P^{-1}\mathfrak{A}P = \mathfrak{C}$ is a complex set.*

If $MA = AM$ where A is any element of $\mathfrak{A}$, then

$$(kI)(P^{-1}AP) = (P^{-1}AP)(kI)$$

and, since k is non-real complex, $P^{-1}AP$ is complex for any A in $\mathfrak{A}$. The converse is immediate.

It is now desirable to separate irreducible semigroups of matrices with quaternion elements into two classes: these which are similar to a complex semigroup under some quaternion similarity transformation, and those which are not similar to a complex semigroup under any such transformation. These cases are considered in turn.

THEOREM 3. *Let $\mathfrak{A}$ be a semigroup of quaternion matrices of degree n which is not similar to a complex set. If $\mathfrak{A}$ is irreducible, then $\mathfrak{A}$ has l-rank n^2; and, conversely, if every semigroup similar to $\mathfrak{A}$ has l-rank n^2, then $\mathfrak{A}$ is irreducible.*

This will be shown in two ways; in the first use is made of a known theorem while in the second a direct proof is given.

Let $\mathfrak{A}$ be a given irreducible semigroup as described. Using the notation and terminology of Brauer (**1**, pp. 513, 520), let $\mathfrak{C}(\mathfrak{A})$ denote the commuting ring of $\mathfrak{A}$, i.e., the set of all quaternion matrices P which intertwine a set $\mathfrak{A}$ of square matrices with itself. This means that for each A of $\mathfrak{A}$, $AP = PA$. It has been seen above that any matrix which commutes with each element of an irreducible semigroup is of the form $P^{-1}(cI)P$. Because of the given nature of $\mathfrak{A}$ c is real. Therefore any $M = P^{-1}(cI)P = cI$ is a real scalar matrix so that here $\mathfrak{C}(\mathfrak{A})$ is the set of all real scalar matrices. Theorem (9.2A) of Brauer's work states the following: let G be an irreducible semigroup of degree n. If G has l-rank k and $\mathfrak{C}(\mathfrak{A})$ has r-rank v, then $n^2 \leqslant kv$. In this instance the r-rank v of $\mathfrak{C}(\mathfrak{A})$ is obviously 1 so that $n^2 \leqslant k$. On the other hand $k \leqslant n^2$, so that $k = n^2$.

This same result can be obtained directly as in the complex case as follows. Let $A = (a_{\kappa\lambda})$ be any quaternion matrix of an irreducible semigroup $\mathfrak{A}$ of order n as given. There may exist n^2 quaternion numbers $k_{\lambda\kappa}$, $\lambda = 1, 2, \ldots, n$; $\kappa = 1, 2, \ldots, n$ such that for each A in $\mathfrak{A}$

$$\sum_{\kappa,\lambda=1}^{n} a_{\kappa\lambda} k_{\lambda\kappa} = 0;$$

if $K = (k_{\lambda\kappa})$ and if

$$\chi(AK) = \sum_{\kappa,\lambda=1}^{n} a_{\kappa\lambda} k_{\lambda\kappa},$$

this can be expressed, as usual, as $\chi(AK) = 0$. If there exists more than one such K-matrix, any right linear combination of them will also be such a K-matrix.

LEMMA. *If $\mathfrak{A}$ has l-rank r where $\mathfrak{A}$ is of degree n, then the number of right linearly independent K-matrices is $n^2 - r$.*

For if

$$A_\rho = (a_{\kappa\lambda}^{(\rho)}), \qquad \rho = 1,2, \ldots, r,$$

is a set of left linearly independent matrices of $\mathfrak{A}$, for a given K-matrix, then

$$\sum_{\lambda,\kappa=1}^{n} a_{\kappa\lambda}^{(\rho)} k_{\lambda\kappa} = 0, \qquad \rho = 1,2, \ldots, r.$$

This is a right system of r homogeneous linear equations in n^2 unknowns, $k_{\lambda\kappa}$. The r rows of the $r \times n^2$ coefficient matrix are left linearly independent and so (see **2**, p. 41, for example) there exist $n^2 - r$ right linearly independent solutions.

Now if $\mathfrak{A}$ is irreducible and not similar to a complex set, it will be shown that there can exist no system of non-zero K matrices. Let us assume the contrary. As in the complex case the following may be noted first of all: For

any A in $\mathfrak{A}$, AK is itself a K-matrix; if $K_1, \ldots, K_m$ is a (right linearly independent) basis for all K-matrices, each A in $\mathfrak{A}$ determines an $m \times m$ matrix $R = (r_{\sigma\rho})$ from

$$AK_\rho = \sum_{\sigma=1}^{m} K_\sigma r_{\sigma\rho}, \qquad \rho = 1,2, \ldots, m;$$

the set $\mathfrak{R}$ of all such matrices determined from a given basis is a semigroup of degree m which is homomorphic to $\mathfrak{A}$; under a change of a K-basis, the set $\mathfrak{R}$ becomes a similar set $P^{-1}\mathfrak{R}P$ where $P = (p_{ij})$ is a non-singular matrix and

$$L_\rho = \sum_{\sigma=1}^{m} K_\sigma p_{\sigma\rho} \qquad \rho = 1,2, \ldots, m,$$

is the new K-basis; and the K-basis may be chosen so that $\mathfrak{R}$ is of the form

$$\begin{bmatrix} \mathfrak{R}_1 & \mathfrak{R}_2 \\ 0 & \mathfrak{R}_4 \end{bmatrix}$$

where $\mathfrak{R}_1$ is irreducible and of degree m_1 where $1 \leqslant m_1 \leqslant m$. If a right linearly independent set $K_1, \ldots, K_m$ existed, then for any A in $\mathfrak{A}$,

$$AK_\rho = \sum_{\sigma=1}^{m_1} K_\sigma r_{\sigma\rho}, \qquad \rho = 1,2, \ldots, m_1$$

(where only $\mathfrak{R}_1$ is utilized). Let

$$K_\rho = (k_{\lambda\mu}^{(\rho)})$$

so that

$$\sum_{\lambda=1}^{n} a_{\kappa\lambda} k_{\lambda\mu}^{(\rho)} = \sum_{\sigma=1}^{m_1} k_{\kappa\mu}^{(\sigma)} r_{\sigma\rho} \qquad \rho = 1,2, \ldots, m_1;\ \kappa = 1,2, \ldots, n;\ \mu = 1,2, \ldots, n.$$

Let

$$P_u = (k_{iu}^{(j)}),\ u = 1,2, \ldots, n;\ i = 1,2, \ldots, n;\ j = 1,2, \ldots, m_1.$$

Then $AP_u = P_u R$ for all A in $\mathfrak{A}$ and all corresponding R in $\mathfrak{R}_1$, and for $u = 1, 2, \ldots, n$. The conditions of Schur's Lemma are met so that either a given $P_u = 0$ or is non-singular. If all $P_u = 0$ for $u = 1, 2, \ldots, n$, then

$$K_1, K_2, \ldots, K_{m_1}$$

are all zero matrices; but this contradicts their linear independence. If not all $P_u = 0$, take any such non-singular P_u so that $A = P_u R P_u^{-1}$; if the proper change of K-basis is made, R can be taken to be the same as A. For each A in $\mathfrak{A}$, $AP_u = P_u A$ and since the conditions of Theorem 1 are met, each non-singular $P_u = Q_u(c_u I)Q_u^{-1}$ where c_u is complex. From Theorem 2, since $\mathfrak{A}$ is not similar to a complex set under any quaternion similarity transformation, $P_u = c_u I$ $(u = 1, 2, \ldots, n)$ where c_u is real. Then it follows that each K_ρ $(\rho = 1, 2, \ldots, m_1 = n)$, is such that each row except the ρth is zero

and this row is of the form $[c_1, c_2, \ldots, c_n]$ for each. For any A of $\mathfrak{A}$ the matrix AK_ρ has for its jth diagonal position the element $a_{j\rho}\, c_j$ $(\rho = 1, 2, \ldots, m_1 = n)$. Now $\chi(AK_\rho) = 0$ $(\rho = 1, 2, \ldots, n)$, so that if A^T denotes the transpose of any A in $\mathfrak{A}$,

$$A^T \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix},$$

and since all c_i are real, the transpose of the above yields $[c_1, c_2, \ldots, c_n] \cdot A = 0$. $[c_1, c_2, \ldots, c_n] = [0, 0, \ldots, 0]$. Now $\mathfrak{A}$ is irreducible so by the corollary to Schur's Lemma, either $P = 0$ or the representation 0 contains $\mathfrak{A}$ as an irreducible component. Since the latter is not possible, each $c_i = 0$; but this means that each $K\rho = 0$ which contradicts the linearly independent character of the $K\rho$.

The converse of the theorem follows directly.

In connection with the latter proof, it can be verified that the following generalization is true: let $\mathfrak{G}$ be an abstract semigroup; let $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3, \ldots$ be a finite number of irreducible semigroups of quaternion matrices of degrees $m_1, m_2, m_3, \ldots$, respectively, which are homomorphic to $\mathfrak{G}$ (relative to matrix multiplication) such that no two semigroups are similar to each other, and such that none is similar to a semigroup of complex matrices. Then there exists no set of non-zero matrices, $K, L, M, \ldots$ such that $\chi(AK) + \chi(BL) + \chi(CM) + \ldots = 0$ simultaneously for all sets of matrices $A, B, C, \ldots$ which correspond to the same element of $\mathfrak{G}$ and belong to $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3, \ldots$, respectively. A proof may be used which parallels that of the complex case and depends on the direct proof of Theorem 3 above. Also, as in this case, one can then state the following: Let $\mathfrak{A}$ and $\mathfrak{B}$ be irreducible semigroups of quaternion matrices which are homomorphic to a semigroup $\mathfrak{G}$ and are not similar to complex semigroups; if the traces of the elements of $\mathfrak{A}$ and $\mathfrak{B}$ which correspond to the same element of $\mathfrak{G}$ are the same, then $\mathfrak{A}$ and $\mathfrak{B}$ are similar. Since $\chi(A) - \chi(B) = 0$ for any A and B of $\mathfrak{A}$ and $\mathfrak{B}$, respectively, which correspond to the same element of $\mathfrak{G}$, the only alternative is that $\mathfrak{A}$ and $\mathfrak{B}$ are similar.

Consider next the case where $\mathfrak{A}$ is an irreducible semigroup of quaternion matrices of degree n and $P\mathfrak{A}P^{-1} = \mathfrak{K}$ is a complex semigroup; let $\overline{\mathfrak{K}}$ denote the set obtained from $\mathfrak{K}$ by taking the ordinary complex conjugate of each matrix of $\mathfrak{K}$. If Q is a complex matrix such that $CQ = Q\bar{C}$ for each C in $\mathfrak{K}$, $\mathfrak{K}$ and $\overline{\mathfrak{K}}$ will be said to be *interjoined* by Q. There may exist such a Q which is non-singular or there may not. It is convenient to consider these cases separately. In this connection the following may be noted:

THEOREM 4. *Let $\mathfrak{A}$ be an irreducible semigroup and let $P^{-1}\mathfrak{A}P = \mathfrak{K}$ be a complex set. If $\mathfrak{K}$ and $\bar{\mathfrak{K}}$ are interjoined by means of a non-singular complex matrix, then every complex semigroup similar to $\mathfrak{A}$ has this property.*

Let $P^{-1}\mathfrak{A}P = \mathfrak{K}_1$ and $Q^{-1}\mathfrak{A}Q = \mathfrak{K}_2$ be complex. Then $\mathfrak{A} = P\mathfrak{K}_1P^{-1} = Q\mathfrak{K}_2Q^{-1}$ or $\mathfrak{K}_1P^{-1}Q = P^{-1}Q\mathfrak{K}_2$; i.e., $P^{-1}Q = M = M_1 + jM_2$ (where M_1 and M_2 are complex matrices) intertwines $\mathfrak{K}_1$ and $\mathfrak{K}_2$. Since $\mathfrak{K}_1(M_1 + jM_2) = (M_1 + jM_2)\mathfrak{K}_2$, $\mathfrak{K}_1M_1 = M_1\mathfrak{K}_2$ and $\bar{\mathfrak{K}}_1M_2 = M_2\mathfrak{K}_2$. Since $\mathfrak{K}_1$, $\mathfrak{K}_2$, and $\bar{\mathfrak{K}}_1$ are irreducible and all matrices involved are complex, M_1 and M_2 are either 0 or non-singular (except that both cannot be 0) and $Q = P(M_1 + jM_2)$. Now assume there exists a complex matrix S such that $SCS^{-1} = \bar{C}$ for each C in $\mathfrak{K}_1$. Then there exists a matrix K in $\mathfrak{K}_2$ such that at least one of $CM_1 = M_1K$ and $\bar{C}M_2 = M_2K$ holds where M_1 and M_2 are fixed. If the former, then $\bar{M}_1\bar{K}\bar{M}_1^{-1} = \bar{C} = SCS^{-1} = SM_1KM_1^{-1}S^{-1}$; if the latter, $\bar{M}_2\bar{K}\bar{M}_2^{-1} = C = S^{-1}\bar{C}S = S^{-1}M_2KM_2^{-1}S$. Since each K in $\mathfrak{K}_2$ can be accounted for in this way; the desired result is obtained.

A result of Brauer's, of use in what follows, states the following: If $\mathfrak{G}$ is an irreducible semigroup of degree n, let E_i denote the row $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with ith component 1; let h be the largest number of indices $u_1, u_2, \ldots, u_h$ with $1 \leqslant u_i \leqslant n$ such that conditions $\sum E_uC_u = 0$, C_u in $\mathfrak{C}(\mathfrak{G})$, u ranging over $u_1, \ldots, u_h$, imply $C_{u_i} = 0$ for all u_i. Then the l-rank of $\mathfrak{G}$ is equal to nh **(1**, pp. 531–532**)**.

If $P^{-1}\mathfrak{A}P = \mathfrak{K}$ is a complex semigroup, if $\mathfrak{A}$ is irreducible, and if M is a matrix in $\mathfrak{C}(\mathfrak{A})$, then $P^{-1}\mathfrak{A}PP^{-1}MP = P^{-1}MPP^{-1}\mathfrak{A}P$ or $\mathfrak{K}N = N\mathfrak{K}$ where $N = P^{-1}MP = M_1 + jM_2$, M_1 and M_2 complex. Then $\mathfrak{K}M_1 = M_1\mathfrak{K}$ and $\bar{\mathfrak{K}}$ and $\mathfrak{K}$ are interjoined by M_2.

(a) If no non-singular complex matrix interjoins $\mathfrak{K}$ and $\bar{\mathfrak{K}}$, then $M_2 = 0$ and M_1 is any complex scalar matrix. Therefore $M = P(k_iI)P^{-1}$ and $\mathfrak{C}(\mathfrak{A})$ consists of all matrices $P(k_iI)P^{-1}$ where P is fixed and k_i is any real or complex number. Let $P = P_1 + jP_2 = (p_{ij})$ where P_1 and P_2 are complex matrices, and let the rank of the $n \times 2n$ matrix $[P_1, P_2]$ be r. Let $u_1, u_2, \ldots, u_r$ denote a set of natural numbers between 1 and n such that the correspondingly numbered rows of $[P_1, P_2]$ are linearly independent. Form the expression $\sum E_uP(k_uI)P^{-1} = 0$ where the summation is over the above $u_1, u_2, \ldots, u_r$ and k_u is any complex number; this is equivalent to $\sum E_uP(k_uI) = 0$. Since E_uP is the uth row of P, this is equivalent to the system of n linear homogeneous equations $T \cdot \alpha = 0$ in r (complex) unknowns,

$$k_{u_1},\ k_{u_2},\ \ldots,\ k_{u_r},$$

where T is the $n \times r$ matrix such that the element in the ith row and jth column is p_{u_jt} and α is a column vector whose transpose is

$$[k_{u_1},\ k_{u_2},\ \ldots,\ k_{u_r}];$$

let $T = T_1 + jT_2$ where T_1 and T_2 are complex. Since the α vector is to be complex, $T \cdot \alpha = 0$ is equivalent to

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \cdot \alpha = 0 .$$

Since the coefficient matrix is of rank r, only the zero solution is possible, i.e., $P(k_u I)P^{-1} = 0$. If any number of rows greater than r is taken a like set of equations results but not all solutions are necessarily zero so that not all $P(k_u I)P^{-1}$ are zero. Therefore, $h = r$.

THEOREM 5. *If $\mathfrak{A}$ is irreducible of degree n, and if $P^{-1}\mathfrak{A}P = \mathfrak{K}$ is complex where no non-singular complex matrix interjoins $\mathfrak{K}$ and $\overline{\mathfrak{K}}$, then $\mathfrak{A}$ has l-rank rn where r is the rank of the matrix $[P_1, P_2]$ where $P = P_1 + jP_2$, P_1 and P_2 complex.*

(b) If $P^{-1}\mathfrak{A}P = \mathfrak{K}$ is interjoined with $\overline{\mathfrak{K}}$ under a non-singular complex matrix, it is convenient to consider separately the cases in which $\mathfrak{K}$ is real and in which $\mathfrak{K}$ is not real.

If $\mathfrak{K}$ is real, then in the above $\mathfrak{K}M_1 = M_1\mathfrak{K}$ and $\mathfrak{K}M_2 = M_2\mathfrak{K}$ so that both M_1 and M_2 are complex scalar matrices. In this case $\mathfrak{C}(\mathfrak{A})$ consists of all matrices of the form $P(k_t + jl_t)IP^{-1}$ (where k_t and l_t are any complex numbers and P is fixed), i.e., of the form $P(q_t I)P^{-1}$ where q_t is any quaternion element. Consider the matrix P^T and let r be its rank (i.e., the number of columns in every maximal set of right linearly independent columns or the number of rows in every maximal set of left linearly independent rows); choose a maximal set of r right linearly independent columns of P^T and let $u_1, u_2, \ldots, u_r$ denote the corresponding column numbers. Form the sum $\sum E_u P(q_u I)P^{-1} = 0$, as before, over this set of u_t; this is equivalent to $\sum E_u P(q_u I) = 0$, or, as before, to the set of equations $T \cdot \alpha = 0$ where T has as its columns the above mentioned set of r right linearly independent columns of P^T and α is now a column vector of r (quaternion) unknown components. Such a system has only the zero solution; and, as before, if any set of u-indices larger in number than r were taken, non-zero solutions could be obtained.

THEOREM 6. *If $\mathfrak{A}$ is irreducible of degree n and if $P^{-1}\mathfrak{A}P = \mathfrak{K}$ is real, then $\mathfrak{A}$ has l-rank rn where r is the rank of P^T.*

If $\mathfrak{K}$ is not real but is interjoined with $\overline{\mathfrak{K}}$ by means of a non-zero complex matrix, then (with reference to the paragraph above (a)) $M_1 = kI$ is complex scalar and an M_2 which is non-singular exists so that $P^{-1}MP = kI + jM_2$, M_2 non-singular. Let S be any other non-singular complex matrix interjoining $\mathfrak{K}$ and $\overline{\mathfrak{K}}$. Then $\bar{C} = SCS^{-1} = M_2 C M_2^{-1}$ for any C in $\mathfrak{K}$ so that $M_2^{-1}SC = CM_2^{-1}S$ for any C in $\mathfrak{K}$; since $\mathfrak{K}$ is irreducible, $M_2^{-1}S = lI$, l complex, so that $S = lM_2$. In this case $\mathfrak{C}(\mathfrak{A})$ consists of all matrices of the form $P(k_t I + jl_t M_2)P^{-1}$ where P and M_2 are fixed and k_t and l_t are complex scalars. (It may be noted that M_2 is itself a non-zero complex scalar if and only if $\mathfrak{K}$ is real). Let $M_2 = Z = (z_{ij})$ for simplicity. Let $P = P_1 + jP_2$, as before, and form $S = [P_1, P_2]$; if t rows of S are chosen to form a matrix S_1, S_1

may be considered to be of the form $S_1 = [Q,R]$ where Q and R are composed of corresponding rows of P_1 and P_2, respectively. Let s be the maximum number of rows of S which are linearly independent and form $[Q,R]$ such that

$$\begin{bmatrix} Q & R \\ -\bar{R}Z & \bar{Q}Z \end{bmatrix}$$

is of rank $2s$. Then consider

$$\sum_u E_u P(k_u I + j l_u Z)P^{-1} = 0$$

where the summation is to be taken over the numbers of the s rows of S chosen as above, $u_1, u_2, \ldots u_s$. After discarding the P^{-1} on the right and noting that Z, k_i and l_i are in the complex field, it can be seen that this relationship is equivalent to the set of $2n$ equations in $2s$ (complex) unknowns

$$\begin{bmatrix} Q^T & -Z^T\bar{R}^T \\ R^T & Z^T\bar{Q}^T \end{bmatrix} \cdot \alpha = 0$$

where

$$\alpha^T = [k_{u_1}, \ldots, k_{u_s}, l_{u_1}, \ldots, l_{u_s}].$$

Since the coefficient matrix is of rank $2s$, only the 0 solution is possible; i.e., if the above relation is to hold, all the matrices from $\mathfrak{C}(\mathfrak{A})$ must be 0 matrices. It is evident that s is the largest number of u_i such that this is true.

THEOREM 7. *If $\mathfrak{A}$ is irreducible of degree n, if $P^{-1}\mathfrak{A}P = \mathfrak{N}$ is non-real complex, and if $\mathfrak{N}$ and $\overline{\mathfrak{N}}$ are interjoined by means of a non-singular complex Z, then $\mathfrak{A}$ has l-rank sn where, when $P = P_1 + jP_2$, s is the maximum number of linearly independent rows of $[P_1,P_2]$ which form $[Q,R]$ such that*

$$\begin{bmatrix} Q & R \\ -\bar{R}Z & \bar{Q}Z \end{bmatrix}$$

is of rank $2s$.

REFERENCES

1. R. Brauer, *On sets of matrices with coefficients in a division ring*, Trans. Amer. Math. Soc., *49* (1941), 502–547.
2. E. H. Moore, *General analysis*, Part I, Memoirs of the American Philosophical Society, I (1935).
3. N. A. Wiegmann, *Some theorems on matrices with real quaternion elements*, Can. J. Math., *7* (1955), 191–201.

Catholic University, Washington, D.C.

PRIME POWER REPRESENTATIONS OF FINITE LINEAR GROUPS II

ROBERT STEINBERG

1. Introduction. The aim of this paper is two-fold: first, to extend the results of **(4)** to the exceptional finite Lie groups recently discovered by Chevalley **(1)**, and, secondly, to give a construction which works simultaneously for the groups A_n, B_n, C_n, D_n, E_n, F_4 and G_2 (in the usual Lie group notation), and which depends only on intrinsic structural properties of these groups. It seems likely that the statements of this paper, especially (1) to (14) upon which the later work is based, also hold for the other known simple linear groups, namely the unitary and second orthogonal groups **(4)**.

Throughout this paper, the phrase "finite Lie group" and the symbol L refer to any of the groups in the first list above. We lean heavily on the properties of these groups developed by Chevalley **(1)**, and use his notations, slightly modified. The symbols u, h, etc., always denote elements of the groups U, H, etc., respectively, these groups being defined in §2.

2. Basic properties of finite Lie groups. In this section we set forth the properties of the groups $G = L$ to be used in the sequel. Statements (1) to (10) are extracted from **(1)**, and the others are proved in §4.

(1) *There exist two subgroups U and H of G such that $U \cap H = 1$, $U H$ is a group, and U is normal in $U H$* **(1**, *p.* 40, Lemma 9**)**.

(2) *There exists a group W (the Weyl group) and for each $w \in W$ an element of G which is also denoted by w such that the union of the sets $H w$ is a group, H is a normal subgroup, and the quotient group is the isomorphic image of W under the map $w \to H w$* **(1**, p. 37 Lemma 3**)**.

(3) *Corresponding to each $w \in W$, U has two subgroups U_w' and U_w* (this is U_w'' in **(1)**) *such that*

(4) $U = U_w' U_w$,

(5) $w U_w' w^{-1} \subseteq U$ *and*

(6) $U_{w_0} = U$ *for some* $w_0 \in W$ **(1**, pp. 41–43**)**.

(7) *G is the union of the sets $U H w U_w$, and*

$$u h w_1 u_{w_1} = \bar{u} \bar{h} w_2 u_{w_2}$$

implies that $u = \bar{u}$, $h = \bar{h}$, $w_1 = w_2$, and

$$u_{w_1} = u_{w_2}$$

(1, p. 42, Theorem 2**)**.

Received November 26, 1956.

(8) *W contains a set of elements $\{w_\alpha\}$ (the fundamental reflections) such that*
(9) *$w_\alpha^2 = 1$ for each α and*
(10) *$\{w_\alpha\}$ generates W* (3, p. 16–05, Theorem 1).
(11) *For each α, if $w = w_\alpha$, we set $U_w = U_\alpha$ and $U_w' = U_\alpha'$; then the union of $U_\alpha H$ and $U_\alpha H w_\alpha U_\alpha$ is a group.*
(12) *For each w and α, at least one of*

$$U_\alpha \subseteq U_w', \; U_\alpha \subseteq U_{ww_\alpha}'$$

holds.
(13) *There is a homomorphism ϵ of W onto the group $\{1, -1\}$ of two elements such that $\epsilon(w_\alpha) = -1$ for each α.*
(14) *There is an element u such that $u \notin U_w'$ for all $w \neq 1$.*

These properties are not independent. For example, the condition $U \cap H = 1$ of (1) follows from (7), and (9) follows from (4), (5), (7) and (11). The complete list is given here for ready reference in what follows.

3. The ideal I and the representation R. In this section, we assume that G is a group for which (1) to (14) hold. Let F be any field and A the group algebra of G over F. For any subset S of G, let the symbol S also denote the sum of the members of S considered as elements of A, and $|S|$ the cardinality of S. Let e be the element of A defined by

$$e = |U H| \Sigma \epsilon(w)\, w, \tag{15}$$

the summation being over the elements of W. Then we can state the following fundamental result.

LEMMA 1. *Let e be defined by* (15). *Then* (i) $e\, w_\alpha = -e$; (ii) *if $u_\alpha \neq 1$, then*

$$w_\alpha u_\alpha w_\alpha^{-1} = \tilde{u}_\alpha h w_\alpha \tilde{u}_\alpha \tag{16}$$

for some $\tilde{u}_\alpha$, $\tilde{u}_\alpha$ and h, and

$$e(u_\alpha w_\alpha^{-1} - \tilde{u}_\alpha + 1) = 0. \tag{17}$$

Proof. Equation (i) follows from (2) and (13). By (7), $w_\alpha u_\alpha \notin U_\alpha H w_\alpha$. Thus $w_\alpha u_\alpha w_\alpha^{-1} \notin U_\alpha H$, and (11) implies (16). By (9) and (12), each right coset of W relative to the group $\{1, w_\alpha\}$ consists of two elements v and vw_α such that $U_\alpha \subseteq U_v'$. Hence, by (1), (2), (5), (9), (13) and (16), we get

$$U H v u_\alpha w_\alpha^{-1} = U H v w_\alpha, \tag{18}$$
$$U H v = U H v \tilde{u}_\alpha \text{ and} \tag{19}$$
$$U H v w_\alpha u_\alpha w_\alpha^{-1} = U H v w_\alpha \tilde{u}_\alpha. \tag{20}$$

Now to get (17), multiply (18), (19), and (20) by $\epsilon(v)$, $\epsilon(v)$ and $\epsilon(vw_\alpha)$, add, and then sum on the right cosets of W relative to the group $\{1, w_\alpha\}$.

THEOREM 1. *Let e be given by* (15), *and let I be the right ideal of A generated by e. Then the set $B = \{e\, u, u \in U\}$ is a vector space basis for I. The dimension of I is $|U|$.*

Proof. For a fixed u, by (4), (5), (6), and (7), the coset $U H w_0 u$ is present in $e u$ with a coefficient $\epsilon(w_0)$ and is not present in $e u_1$ if $u_1 \neq u$. Thus B is linearly independent. Let I_0 be the linear span of B. Then $e \in I_0 \subseteq e A = I$. Thus $I_0 = I$ if it can be shown that I_0 is an ideal. For any u and h, $I_0 u \subseteq I_0$ and $I_0 h \subseteq I_0$, by (1) and (2). Also for any u and α, by (4) we can write $u = u_\alpha u_\alpha'$ and then $e u w_\alpha^{-1} = (e u_\alpha w_\alpha^{-1})(w_\alpha u_\alpha' w_\alpha^{-1}) \in I_0$, using Lemma 1 and (5); hence $I_0 w_\alpha^{-1} \subseteq I_0$. By (2), (7), and (10), A is generated by elements of the form u, h and w_α^{-1}. Thus $I_0 A \subseteq I_0$ and Theorem 1 is proved.

COROLLARY. *If $s \in I$ and $s = \Sigma \gamma_u e u$, $\gamma_u \in F$, then γ_u is $\epsilon(w_0)$ times the coefficient of $U H w_0 u$ in the right coset decomposition of s relative to $U H$, and $\Sigma \gamma_u$ is the coefficient of $U H$. If $x \in G$ and $e x = \Sigma \gamma_u e u$, then $\gamma_u = 1, -1$ or 0, and at most $|W|$ of the γ_u are non-zero.*

Proof. The first statement follows from (6), (7) and Theorem 1. Since e is the signed sum of $|W|$ right cosets relative to $U H$, the same is true of $e x$, and the last statement now follows from the first.

LEMMA 2. *Let m be the index of $U H$ in G: $m = \Sigma |U_w|$. Then*

$$e U \Sigma \epsilon(w) w = m e. \tag{21}$$

Proof. For each $x \in G$, write $e x = \Sigma \gamma(x; u) e u$ so that

$$e = \Sigma \gamma(x; u) e u x^{-1}. \tag{22}$$

Here $\gamma(x; u)$ is $\epsilon(w w_0)$ if

$$U H w_0 u = U H w x \tag{23}$$

for some w and is 0 otherwise, by the corollary to Theorem 1. Now (23) is equivalent to $u x^{-1} \in w_0^{-1} U H w$. Thus summation of (22) on $x \in G$ gives

$$|G| e = \Sigma_u \Sigma_w \epsilon(w w_0) e w_0^{-1} U H w = \Sigma_w e U H \Sigma_w \epsilon(w) w = |U H| e U \Sigma \epsilon(w) w$$

by (1), (2) and (13). If the base field F has characteristic 0, division by $|U H|$ yields (21). Since only integral coefficients occur in (21), it remains valid for any field.

We now state the principal result of this section.

THEOREM 2. *Let G be a group for which (1) to (13) hold. Let I and B be as in Theorem 1, let m be the index of $U H$ in G, and let R be the representation of G on I by right multiplication. Then*

(i) *R restricted to U is equivalent to the right regular representation of U; the degree of R is $|U|$;*

(ii) *relative to B, R is represented by a set of matrices each of which has only entries of $1, -1$ or 0 with at most $|W|$ non-zero entries in each row;*

(iii) *if the base field F has characteristic 0 or prime to m, then R is irreducible.*

Proof. Theorem 1 and its corollary imply (i) and (ii). We prove (iii) by showing that any element $s \neq 0$ of I generates all of I under right multiplica-

tion by the elements of A, the linear span of G. If $s \neq 0$, there is a coset $U H x$ represented in s with a coefficient $\gamma \neq 0$. Then $U H$ has the coefficient γ in $s x^{-1}$. By the corollary to Theorem 1, $s x^{-1} = \Sigma \gamma_u e u$ with $\Sigma \gamma_u = \gamma$. Hence, by (1) and Lemma 2,

$$s x^{-1} (\gamma m)^{-1} U \Sigma \epsilon(w) w = (\gamma m)^{-1} \Sigma_u \gamma_u e u U \Sigma_w \epsilon(w) w = e.$$

Thus $I = e A \subseteq s A \subseteq I$, and (iii) is proved.

We use the condition (14) for the first time in the proof of the following converse to Theorem 2.

THEOREM 3. *In the notation of Theorem* 2, *if* (1) *to* (14) *hold for G and if the characteristic of F divides m, then R is reducible.*

Proof. By (6) and (7), $e U \neq 0$ so that $e U$ generates a non-zero invariant subspace I_1 of I. We complete the proof by showing that $I_1 \neq I$ or, equivalently, that $e \notin I_1$. Lemma 2, with the elements of G replaced by their inverses, implies, because of (1), (2), and (13), that $\Sigma \epsilon(w) w$ is a left annihilator of $e U$ and hence of I_1. But it is not a left annihilator of e. Indeed we show that an element u for which (14) holds appears in $\Sigma \epsilon(w) w e$ with a coefficient 1. Now $u \in w_1 U H w_2$ implies that $w_1^{-1} u \in U H w_2$. By (2) (7), and (14), $w_1 = 1$ and then $w_2 = 1$. Again by (2) and (7), u can be written uniquely as $1 \bar{u} \bar{h} 1$, so that all assertions are proved.

4. Representations of finite Lie groups. In order to apply the results of §3 to the groups L, we first prove:

LEMMA 3. *Each group L has properties* (1) *to* (14).

Proof. As has already been stated in §2, (1) to (10) are extracted from **(1)**. In the proof of (11) to (14), we use the standard Lie algebra terminology **(2)**, and note that the index set $\{\alpha\}$ of (8) refers to a simple system of positive roots and the w_α are the reflections in these roots. It is proved in **(3**, p. 19–01, Lemma 1**)** that w_α maps α upon $-\alpha$ and permutes the remaining positive roots. Then (11) follows by **(1**, pp. 36, 41**)**. Also

$$(w w_\alpha)(\alpha) = w(-\alpha) = -w(\alpha).$$

Thus either $(w w_\alpha)(\alpha)$ or $w(\alpha)$ is positive, and (12) is true. For each w, let $n(w)$ be the number of roots r such that $r > 0$ and $w(r) < 0$. Then $\epsilon(w) = (-1)^{n(w)}$ fulfils the requirements of (13). Finally, if $w \neq 1$, then $w(\alpha) < 0$ for at least one α **(3**, p. 16–08, Theorem 2**)**, and so, by the same reasoning as above, $U_w' \subseteq U_\alpha'$ for some α, and similarly, $U_\beta \subseteq U_\gamma'$ if $\beta \neq \gamma$. Thus, if we choose $u_\alpha \neq 1$ for each α and set $u = \Pi u_\alpha$, then $u \notin U_\alpha'$ for each α, hence $u \notin U_w'$ for each $w \neq 1$, and (14) is proved.

If the group L is defined over a field of q elements and of characteristic p, it is further proved in **(1)** that U can be chosen as a p-Sylow subgroup of L,

$U H$ as the normalizer of U, and then $| U_w | = q^{n(w)}$, with $n(w)$ as above. Thus we have:

THEOREM 4. *Let L be a finite Lie group over a field of q elements and of characteristic p. Let U be a p-Sylow subgroup of L, and let m be the index of the normalizer of U in L, this number being given by $m = \Sigma q^{n(w)}$. Let I be any vector space of dimension $| U |$ over a field of characteristic 0 or prime to m. Then L has an irreducible representation R of degree $| U |$ which has I as the representation space and for which the results of Theorem 2 are valid.*

REFERENCES

1. C. Chevalley, *Sur certains groupes simples*, Tohoku Math. J., 7 (1955), 14–66.
2. E. B. Dynkin, *The structure of semi-simple algebras*, Amer. Math. Soc. Translation No. 17.
3. Séminaire "Sophus Lie" (Paris, 1954–5).
4. R. Steinberg, *Prime power representations of finite linear groups*, Can. J. Math., *8* (1956), 580–591.

University of California, Los Angeles

CERTAIN CLASSES OF IDEALS IN POLYNOMIAL RINGS

N. H. McCOY

1. Introduction. The purpose of this paper is to establish some results, of a somewhat miscellaneous nature, concerning certain classes of ideals in polynomial rings. Although the results will be formulated for a ring $R[x]$ of polynomials in one indeterminate over a given ring R, they can be easily extended to polynomial rings in any finite number of indeterminates. We shall be concerned both with right ideals and two-sided ideals, but the word *ideal* used alone will henceforth mean two-sided ideal.

A right ideal $\mathfrak{q}$ in the ring R is said to be a *semi-prime right ideal* if $\mathfrak{a}^2 \subseteq \mathfrak{q}$ implies that $\mathfrak{a} \subseteq \mathfrak{q}$, where $\mathfrak{a}$ is any right ideal of R. Of course, a semi-prime right ideal which is an ideal will be called a *semi-prime ideal*.

Each class of ideals which we shall consider is a sub-class of the class of semi-prime right ideals. Clearly the class of semi-prime ideals (right ideals) of R is closed under arbitrary intersection. The unique minimal semi-prime ideal of R is the *lower radical* of R as defined by Baer **(2)**. If the lower radical of R is zero, we call R a *semi-prime ring*. Thus a ring R is semi-prime if and only if the zero ideal is a semi-prime ideal of R. It is known **(1, 10)** that if N is the lower radical of R, then the lower radical of $R[x]$ is $N[x]$. New proofs of this fact will appear in §3 and §4. In particular, the polynomial ring $R[x]$ is semi-prime if and only if R is semi-prime.

Johnson **(7)** has introduced the concept of *prime right ideal* in a semi-prime ring, and has shown that these ideals play an important role in the structure theory of such rings. The definition will be found in §5. However, we may now state the various classes of ideals which will be considered in this paper. If R is an arbitrary ring, we shall study in R and in $R[x]$:

(1) semi-prime right ideals,

(2) semi-prime ideals, and

(3) prime ideals.

Also, for the case in which R (and therefore also $R[x]$) is semi-prime, we shall consider

(4) prime right ideals, and

(5) annihilating ideals.

All of these are special instances of semi-prime right ideals.

The following terminology will be convenient. Let σ be a property of ideals (right ideals) defined for ideals in R and in $R[x]$. For example, σ may be the property defining any one of the above five classes of ideals. An ideal (right

Received August 15, 1956.

ideal) with property σ may be called a σ-ideal. If σ is such that $\mathfrak{a}$ a σ-ideal in R implies that $\mathfrak{a}[x]$ is a σ-ideal in $R[x]$, we shall find it convenient to say that the "going up" theorem holds for σ-ideals. If A a σ-ideal in $R[x]$ implies that $A \cap R$ is a σ-ideal in R, we shall say that the "going down" theorem holds for σ-ideals. It will be shown that both of these theorems hold for σ-ideals, where σ-ideals are any one of the five classes of ideals mentioned above. This implies a certain closure property which will be obtained in §2.

Let A be a right ideal in $R[x]$ and $\mathfrak{a}$ the right ideal in R generated by all coefficients of elements of A. If $\mathfrak{q}$ is a semi-prime right ideal of R, we shall show in Theorem 4 that $(\mathfrak{q}[x]:A) = (\mathfrak{q}:\mathfrak{a})[x]$. As a special case of this result it follows that if R is a semi-prime ring, the only annihilating ideals in $R[x]$ are those ideals of the form $\mathfrak{a}[x]$, where $\mathfrak{a}$ is an annihilating ideal in R. As a consequence of this fact, we can say something about the structure of $R[x]$ as it is induced by the structure of R. This is based on results of Johnson **(7)** and will be discussed in §6. In the final section we make a few remarks about prime right ideals in $R[x]$, where R is a *division ring*, and give an example that may be of some interest.

2. A general closure property. Before considering any particular class of ideals, we prove a result in a fairly general setting.

Let σ be a property of ideals (right ideals) defined in the ring R and also in the polynomial ring $R[x]$. An ideal (right ideal) with property σ may be called a σ-ideal. As an example, σ-ideals may be the semi-prime right ideals. If R is itself a σ-ideal, the property σ may be used to define a closure operation on the set of all ideals (right ideals) of R. If $\mathfrak{a}$ is an ideal (right ideal) in R, the *closure* of $\mathfrak{a}$ relative to σ, which we may denote by $\mathfrak{a}_\sigma$, is the intersection of all σ-ideals which contain $\mathfrak{a}$. Of course, the *closed ideals* (relative to σ) are those ideals that are intersections of σ-ideals.

We shall prove the following result:

THEOREM 1. *Let σ be a property of ideals (right ideals) defined in R and in $R[x]$ such that the following are true:*

(a) *R is a σ-ideal,*

(b) *If A is an ideal (right ideal) and B a σ-ideal in $R[x]$ such that $AR[x] \subseteq B$, then $A \subseteq B$,*

(c) *The "going up" and "going down" theorems hold for σ-ideals.*

Then for each ideal (right ideal) $\mathfrak{a}$ in R, $(\mathfrak{a}[x])_\sigma = \mathfrak{a}_\sigma[x]$.

Clearly (a) and (c) imply that $R[x]$ is a σ-ideal in $R[x]$. Moreover, it follows easily from (c) that the "going up" and "going down" theorems hold also for *closed* ideals.

Since $\mathfrak{a} \subseteq \mathfrak{a}_\sigma$, we have $\mathfrak{a}[x] \subseteq \mathfrak{a}_\sigma[x]$. But by the "going up" theorem for closed ideals, $\mathfrak{a}_\sigma[x]$ is a closed ideal in $R[x]$ and hence $(\mathfrak{a}[x])_\sigma \subseteq \mathfrak{a}_\sigma[x]$. To obtain inclusion in the other direction, suppose that $\mathfrak{a}[x] \subseteq C$, where C is any closed ideal in $R[x]$. Then $C \cap R$ is a closed ideal in R by the "going

down" theorem, and since $\mathfrak{a} \subseteq C \cap R$ we have $\mathfrak{a}_\sigma \subseteq C \cap R$. It follows that $\mathfrak{a}_\sigma[x]R[x] \subseteq C$ and (b) implies that $\mathfrak{a}_\sigma[x] \subseteq C$. Applying this result to the case in which $C = (\mathfrak{a}[x])_\sigma$, we have $\mathfrak{a}_\sigma[x] \subseteq (\mathfrak{a}[x])_\sigma$ and the proof is completed.

If σ-ideals are any one of the various classes of ideals to be studied in this paper, it will be obvious that (a) and (b) hold. Accordingly, in order to apply this theorem it will only be necessary to establish the "going up" and "going down" theorems.

3. The prime radical of a polynominal ring. In this section we give a simple application of the preceding theorem to the case in which σ-ideal means prime ideal. We need therefore to establish the following result:

LEMMA 1. *The "going up" and "going down" theorems hold for prime ideals.*

The "going up" theorem follows easily from the fact that if $\mathfrak{p}$ is an ideal in R, then $R[x]/\mathfrak{p}[x] \cong (R/\mathfrak{p})[x]$, and $(R/\mathfrak{p})[x]$ is a prime ring (that is, the zero ideal is prime) if and only if $R/\mathfrak{p}$ is a prime ring.

The "going down" theorem has been established in **(10)** but for completeness we reproduce the proof here. If P is a prime ideal in $R[x]$, we wish to prove that $P \cap R$ is a prime ideal in R. By Theorem 1 of **(9)** we only need to show that if a and b are elements of R such that $aRb \subseteq P \cap R$, then $a \in P \cap R$ or $b \in P \cap R$. If $aRb \subseteq P \cap R$, then each element of $aR[x]bR[x]$ is a sum of terms belonging to P and thus $aR[x]bR[x] \subseteq P$. Since P is a prime ideal in $R[x]$ it follows that $a \in P$ or $b \in P$, and hence $a \in P \cap R$ or $b \in P \cap R$, as required.

By the *prime radical* of the ring R we shall mean the radical as defined in **(9)**, that is, the intersection of all prime ideals in R. It is known **(8; 12)** that the prime radical coincides with the *lower radical* of Baer **(2)**. If σ-ideal means prime ideal, then the prime radical of R is the closure of the zero ideal in the terminology of the preceding section. A simple application of Theorem 1 then yields the following result:

THEOREM 2. *If K is the prime radical of the ring R, the prime radical of the polynomial ring $R[x]$ is $K[x]$.*

Different proofs of this result are to be found in **(1)** and **(10)**.

4. Semi-prime right ideals. Before passing to a consideration of polynomial rings we make a few preliminary remarks about semi-prime right ideals in general.

A right ideal $\mathfrak{q}$ in the ring R is said to be *semi-prime* if $\mathfrak{a}^2 \subseteq \mathfrak{q}$ implies that $\mathfrak{a} \subseteq \mathfrak{q}$, where $\mathfrak{a}$ is any right ideal in R. Henceforth we shall find it convenient to denote the sets of right ideals, ideals, semi-prime right ideals, and semi-prime ideals of any ring R by $\mathfrak{I}_r(R)$, $\mathfrak{I}(R)$, $\mathfrak{S}_r(R)$, and $\mathfrak{S}(R)$, respectively.

(A) *If $\mathfrak{q} \in \mathfrak{I}_r(R)$, then $\mathfrak{q} \in \mathfrak{S}_r(R)$ if and only if $a \in R$ such that $aRa \subseteq \mathfrak{q}$ implies that $a \in \mathfrak{q}$.*

To show this, suppose that $\mathfrak{q} \in \mathfrak{S}_r(R)$ and that $aRa \subseteq \mathfrak{q}$. Then $(aR)^2 \subseteq \mathfrak{q}$ and this implies that $aR \subseteq \mathfrak{q}$. If $(a)_r$ denotes the right ideal of R generated by a, we now have $[(a)_r]^2 \subseteq \mathfrak{q}$ and hence $a \in \mathfrak{q}$. Conversely, let $\mathfrak{a} \in \mathfrak{J}_r(R)$ such that $aRa \subseteq \mathfrak{q}$ implies that $a \in \mathfrak{q}$, and let $\mathfrak{a}$ be a right ideal with $\mathfrak{a}^2 \subseteq \mathfrak{q}$. If $a \in \mathfrak{a}$, then $aRa \subseteq \mathfrak{a}^2 \subseteq \mathfrak{q}$ and $a \in \mathfrak{q}$. Hence $\mathfrak{a} \subseteq \mathfrak{q}$ and $\mathfrak{q} \in \mathfrak{S}_r(R)$.

(B) *If* $\mathfrak{q} \in \mathfrak{S}_r(R)$ *and* $a, b \in R$ *such that* $aRb \subseteq \mathfrak{q}$, *then* $ab \in \mathfrak{q}$.

Clearly $aRb \subseteq \mathfrak{q}$ implies that $(abR)^2 \subseteq \mathfrak{q}$. It follows that $abR \subseteq \mathfrak{q}$, and the desired result follows from (A).

We may remark in passing that Herz **(3)** has shown that if R is commutative, the lattice $\mathfrak{S}(R)$ is distributive. More recently, Tominaga **(13)** has established the same result for an arbitrary ring R.

We shall next prove the following useful lemma.

LEMMA 2. *Let* $\mathfrak{q}$ *be a semi-prime right ideal in the arbitrary ring* R, *and let* f *and* g *be elements of the polynomial ring* $R[x]$ *such that* $fRg \subseteq \mathfrak{q}[x]$. *Then* $aRb \subseteq \mathfrak{q}$, *where* a *is any coefficient of* f *and* b *is any coefficient of* g.

This is obvious if g is of degree zero, and we proceed to use induction on the degree of g. Let us set

$$f = a_0x^n + \ldots + a_n, \qquad (a_0 \neq 0),$$

and

$$g = b_0x^m + \ldots + b_m, \qquad (b_0 \neq 0).$$

Hence we have

$$(1) \qquad (a_0x^n + \ldots + a_n)R(b_0x^m + \ldots + b_m) \subseteq \mathfrak{q}[x].$$

By considering the coefficients of $x^{n+m}, \ldots, x^m$, we obtain the following in which r is an arbitrary element of R:

$$(2) \qquad \begin{aligned} &a_0rb_0 \in \mathfrak{q}, \\ &(a_0rb_1 + a_1rb_0) \in \mathfrak{q}, \\ &(a_0rb_2 + a_1rb_1 + a_2rb_0) \in \mathfrak{q}, \\ &\qquad \ldots \\ &(\quad \ldots + a_nrb_0) \in \mathfrak{q}. \end{aligned}$$

If we multiply the second of these relations by sb_0 on the right, where s is an arbitrary element of R, we get

$$(3) \qquad (a_0rb_1sb_0 + a_1rb_0sb_0) \in \mathfrak{q},$$

and the first term on the left side of (3) is in $\mathfrak{q}$ by the first of relations (2). Hence $a_1Rb_0Rb_0 \subseteq \mathfrak{q}$, and it follows that $(a_1Rb_0R)^2 \subseteq \mathfrak{q}$. Since $\mathfrak{q}$ is semi-prime, this implies that $a_1Rb_0R \subseteq \mathfrak{q}$ and then property (A) shows that $a_1Rb_0 \subseteq \mathfrak{q}$.

Now if we multiply the third of relations (2) by sb_0 on the right, the first two terms are in $\mathfrak{q}$ and it follows that $a_2rb_0sb_0 \in \mathfrak{q}$. As above, this implies that $a_2Rb_0 \subseteq \mathfrak{q}$. Obviously, this can be continued to yield

$$a_iRb_0 \subseteq \mathfrak{q} \ (i = 0,1, \ldots, n).$$

Thus $fRb_0 \subseteq \mathfrak{q}[x]$ and hence from (2) we have

$$fR(b_1x^{m-1} + \ldots + b_m) \subseteq \mathfrak{q}[x],$$

and the induction hypothesis shows that $aRb \subseteq \mathfrak{q}$, where a and b are any coefficients of f and g, respectively.

We shall now prove the following theorem.

THEOREM 3. *The "going up" and "going down" theorems hold for semi-prime right ideals.*

Suppose, first, that $\mathfrak{q} \in \mathfrak{S}_r(R)$ and let us show that $\mathfrak{q}[x] \in \mathfrak{S}_r(R[x])$. By (A), we only need to show that if $f \in R[x]$ such that $fR[x]f \subseteq \mathfrak{q}[x]$, then $f \in \mathfrak{q}[x]$. In particular, $fRf \subseteq \mathfrak{q}[x]$ and Lemma 2 shows that $aRb \subseteq \mathfrak{q}$, where a and b are any coefficients of f. Since also $ab \in \mathfrak{q}$ by property (B) it follows that $\mathfrak{a}^2 \subseteq \mathfrak{q}$, where $\mathfrak{a}$ is the right ideal in R generated by the coefficients of f. Hence $\mathfrak{a} \subseteq \mathfrak{q}$ and $f \in \mathfrak{q}[x]$.

To prove the "going down" theorem, suppose that $Q \in \mathfrak{S}_r(R[x])$ and let us show that $\mathfrak{q} = Q \cap R \in \mathfrak{S}_r(R)$. Clearly $(\mathfrak{q}[x])^2 \subseteq Q$, and therefore $\mathfrak{q}[x] \subseteq Q$. If $a \in R$ such that $aRa \subseteq \mathfrak{q}$, it follows that

$$(aR[x])^2 \subseteq \mathfrak{q}[x] \subseteq Q$$

and hence $aR[x] \subseteq Q$. However, by (A), we then have $a \in Q$ so that $a \in \mathfrak{q}$, completing the proof.

Remark. Although we have stated and proved this result for semi-prime right ideals, it is clear that it holds also for semi-prime *ideals*. In this case, the proof of the "going up" theorem need not depend on Lemma 2 but can be easily established as follows. Let $fR[x]f \subseteq \mathfrak{q}[x]$, where f is of degree n, and let a be the leading coefficient of f. It follows that $aRa \subseteq \mathfrak{q}$ and hence $a \in \mathfrak{q}$. Therefore $ax^n \in \mathfrak{q}[x]$ and we have

$$(f - ax^n)R[x](f - ax^n) \subseteq \mathfrak{q}[x].$$

Since $f - ax^n$ is zero or has degree less than n, it is clear how to complete the proof by induction on n.

From Theorems 1 and 3 we now obtain at once the following result.

COROLLARY. *If σ-ideals are taken to be the semi-prime right ideals and $\mathfrak{a}$ is any right ideal in R, then* $(\mathfrak{a}[x])_\sigma = \mathfrak{a}_\sigma[x]$.

We may also point out that if σ-ideals are taken to be the semi-prime *ideals*, the lower radical of R is just $(0)_\sigma$ and it follows that if N is the lower radical of R, then the lower radical of $R[x]$ is $N[x]$. Since the lower radical coincides with the prime radical, this is still another proof of Theorem 2.

If $\mathfrak{a}, \mathfrak{b} \in \mathfrak{I}_r(R)$, we shall use the familiar notation

$$(\mathfrak{b}:\mathfrak{a}) = \{r; r \in R, \mathfrak{a}r \subseteq \mathfrak{b}\}.$$

In case $\mathfrak{b} = 0$, we shall write $\mathfrak{a}^r$ in place of $((0):\mathfrak{a})$. Hence $\mathfrak{a}^r$ is the ideal of right annihilators of the right ideal $\mathfrak{a}$. We next prove the following theorem.

THEOREM 4. *Let $A \in \mathfrak{I}_r(R[x])$ and $\mathfrak{q} \in \mathfrak{S}_r(R)$. Then if $\mathfrak{a}$ is the right ideal in R generated by all coefficients of elements of A, we have $(\mathfrak{q}[x]:A) = (\mathfrak{q}:\mathfrak{a})[x]$.*

Let f be any element of $(\mathfrak{q}:\mathfrak{a})[x]$, and c any coefficient of f. Then $\mathfrak{a}c \subseteq \mathfrak{q}$, which implies that $Ac \subseteq \mathfrak{q}[x]$ and hence that $Af \subseteq \mathfrak{q}[x]$. It follows that $(\mathfrak{q}:\mathfrak{a})[x] \subseteq (\mathfrak{q}[x]:A)$. This part is trivial and makes no use of the hypothesis that q is semi-prime.

Now assume that $f \in (\mathfrak{q}[x]:A)$. If a is any coefficient of an element of A and c is any coefficient of f, Lemma 2 assures us that $aRc \subseteq \mathfrak{q}$. Since this implies that $ac \in \mathfrak{q}$, we have $\mathfrak{a}c \subseteq \mathfrak{q}$ or $c \in (\mathfrak{q}:\mathfrak{a})$. Thus $f \in (\mathfrak{q}:\mathfrak{a})[x]$, and hence $(\mathfrak{q}[x]:A) \subseteq (\mathfrak{q}:\mathfrak{a})[x]$. The proof is therefore completed.

We recall that a semi-prime *ring* is a ring in which the zero ideal is semi-prime. The following result is then obtained from the theorem just proved by specifying q to be the zero ideal.

COROLLARY. *If A is a right ideal in the semi-prime ring $R[x]$ and $\mathfrak{a}$ is the right ideal in R generated by all coefficients of elements of A, then $A^r = \mathfrak{a}^r[x]$.*

Thus, if R is a semi-prime ring, $Ag = 0$ if and only if each coefficient of g is a right annihilator of all coefficients of elements of A. This is not true if R is an arbitrary ring, as will be apparent from an example to be given below, but it is known **(11)** that $A^r \neq 0$ if and only if $\mathfrak{a}^r \neq 0$.

THEOREM 5. *Suppose that $\mathfrak{q}[x] \subseteq A$, where $\mathfrak{q} \in \mathfrak{S}_r(R)$ and $A \in \mathfrak{I}_r(R[x])$, R being an arbitrary ring. If $\mathfrak{a}$ is the right ideal in R generated by all coefficients of elements of A and $\mathfrak{b}$ is the right ideal in R generated by the leading coefficients of elements of A, then $(\mathfrak{q}:\mathfrak{a}) = (\mathfrak{q}:\mathfrak{b})$.*

Since $\mathfrak{b} \subseteq \mathfrak{a}$, it is clear that $(\mathfrak{q}:\mathfrak{a}) \subseteq (\mathfrak{q}:\mathfrak{b})$. Let $c \in (\mathfrak{q}:\mathfrak{b})$, and let us show that $c \in (\mathfrak{q}:\mathfrak{a})$. That is, if $f \in A$, we wish to show that

$$fc \in \mathfrak{q}[x]. \tag{4}$$

This is true if f is of degree zero, for then $f \in \mathfrak{b}$. Let us therefore set

$$f = a_0x^n + \ldots + a_n \qquad (a_0 \neq 0) \tag{5}$$

and assume (4) for all elements of A of degree less than n. Clearly $a_0 \in \mathfrak{b}$ and $a_0R \subseteq \mathfrak{b}$, so that $a_0Rc \subseteq \mathfrak{q}$ and therefore $a_0c \in \mathfrak{q}$. Hence $a_0cx^n \in \mathfrak{q}[x] \subseteq A$, and it follows that

$$(a_1x^{n-1} + \ldots + a_n)c \subseteq A.$$

By use of the induction hypothesis we can conclude that $a_icRa_ic \subseteq \mathfrak{q}$, and hence finally that $a_ic \in \mathfrak{q}$ $(i = 0, 1, \ldots, n)$. It follows that $fc \in \mathfrak{q}[x]$, and the proof is completed.

In case R is a semi-prime ring, we can choose $\mathfrak{q}$ to be the zero ideal and obtain the following corollary.

COROLLARY. *Let $A \in \mathfrak{I}_r(R)$, where R is a semi-prime ring, and let $\mathfrak{a}$ and $\mathfrak{b}$ be as defined in the preceding theorem. Then $\mathfrak{a}^r = \mathfrak{b}^r$.*

As an example to show that the conclusions of the preceding corollary, and also of this one, are not true for arbitrary rings, let R be the ring $I[t]/(4, t^2)$, where I is the ring of integers and t an indeterminate. In $R[x]$, let $A = (tx + 2)$. Then it is obvious that $(tx + 2) \subseteq A^r$, whereas $t \notin \mathfrak{a}^r$ since $2t \neq 0$. Moreover, it can be verified by a detailed but straight-forward calculation that t annihilates the leading coefficient of each element of A, and hence that $t \in \mathfrak{b}^r$. In particular, then, we have $\mathfrak{a}^r \neq \mathfrak{b}^r$.

5. Prime right ideals in semi-prime rings. Throughout this section and the next we shall consider semi-prime rings only.

A right ideal $\mathfrak{p}$ in a semi-prime ring R is said to be a *prime right ideal* if $\mathfrak{ab} \subseteq \mathfrak{p}$, where $\mathfrak{a}, \mathfrak{b} \in \mathfrak{I}_r(R)$ with $\mathfrak{b}^r = 0$, implies that $\mathfrak{a} \subseteq \mathfrak{p}$.

Johnson **(7)** has introduced this definition and has shown that the prime right ideals play an important role in the structure theory of semi-prime rings. We may emphasize that the concept of prime right ideal has meaning only for semi-prime rings.

By Theorem 2 we know that $R[x]$ is a semi-prime ring if and only if R is semi-prime. It is known **(7**, p. 378**)** that a prime right ideal is semi-prime, and therefore some of our previous results are applicable to prime right ideals. We may denote the set of prime right ideals of R by $\mathfrak{P}_r(R)$, and hence $\mathfrak{P}_r(R) \subseteq \mathfrak{S}_r(R)$ for any semi-prime ring R.

We now prove the following theorem.

THEOREM 6. *The "going up" and "going down" theorems hold for prime right ideals.*

Suppose that $\mathfrak{p} \in \mathfrak{P}_r(R)$ and that $AB \subseteq \mathfrak{p}[x]$, where $A, B \in \mathfrak{I}_r(R[x])$ with $B^r = 0$. If $\mathfrak{a}$ and $\mathfrak{b}$ are the respective right ideals in R generated by the coefficients of elements of A and of B, it follows from Lemma 2 that $\mathfrak{ab} \subseteq \mathfrak{p}$. Moreover, $\mathfrak{b}^r = 0$ and it follows that $\mathfrak{a} \subseteq \mathfrak{p}$ and therefore that $A \subseteq \mathfrak{p}[x]$. This proves the "going up" theorem.

Now let $P \in \mathfrak{P}_r(R[x])$, and set $\mathfrak{p} = P \cap R$. Suppose that $\mathfrak{ab} \subseteq \mathfrak{p}$, where $\mathfrak{a}, \mathfrak{b} \in \mathfrak{I}_r(R)$ with $\mathfrak{b}^r = 0$. Then $\mathfrak{a}[x]\mathfrak{b}[x] \subseteq \mathfrak{p}[x]$ and the Corollary to Theorem 4 implies that $(\mathfrak{b}[x])^r = 0$. But $\mathfrak{p}[x]R[x] \subseteq P$, and this implies that $\mathfrak{p}[x] \subseteq P$. Hence we have $\mathfrak{a}[x]\mathfrak{b}[x] \subseteq P$ with $(\mathfrak{b}[x])^r = 0$, and hence $\mathfrak{a}[x] \subseteq P$. It follows that $\mathfrak{a} \subseteq \mathfrak{p}$, and this establishes the "going down" theorem.

Now let σ-ideal mean prime right ideal. Evidently the set of prime right ideals is closed under arbitrary intersection and if $\mathfrak{a} \in \mathfrak{I}_r(R)$, $\mathfrak{a}_\sigma$ is the unique least prime right ideal which contains $\mathfrak{a}$. In the terminology of **(7)**, $\mathfrak{a}_\sigma$ is the

prime cover of $\mathfrak{a}$. In view of Theorem 6, we may apply Theorem 1 to show that if $\mathfrak{p}$ is the prime cover of $\mathfrak{a}$ in R, then $\mathfrak{p}[x]$ is the prime cover of $\mathfrak{a}[x]$ in $R[x]$.

If $\mathfrak{a} \in \mathfrak{I}_r(R)$, Theorem 6 shows that $\mathfrak{a}[x] \in \mathfrak{P}_r(R[x])$ if and only if $\mathfrak{a} \in \mathfrak{P}_r(R)$. Of course, this says nothing about the prime right ideals of $R[x]$ which are not of the form $\mathfrak{a}[x]$. An example to be given in §7 will show that there may exist many prime right ideals not of this form.

6. Annihilating ideals in semi-prime rings. If T is a set of elements of the semi-prime ring R, we may set $T^r = \{a;\ a \in R,\ Ta = 0\}$ and $T^l = \{a;\ a \in R,\ aT = 0\}$. The right ideal $\mathfrak{a}$ in R is an *annihilating right ideal* if $\mathfrak{a} = T^r$ for some set T. Annihilating left ideals are defined in an analogous way, and an ideal which is both an annihilating right ideal and an annihilating left ideal will be called an *annihilating ideal*. We may denote the set of all annihilating ideals in R by $\mathfrak{A}(R)$. It is known **(7**, p. 376**)** that if $\mathfrak{a} \in \mathfrak{I}(R)$, then $\mathfrak{a} \in \mathfrak{A}(R)$ if and only if $\mathfrak{a} = \mathfrak{a}^{rl}$, where $\mathfrak{a}^{rl}$ means $(\mathfrak{a}^r)^l$. Moreover, it can be shown **(7**, p. 379**)** that $\mathfrak{A}(R) = \mathfrak{I}(R) \cap \mathfrak{P}_r(R)$, and clearly $\mathfrak{A}(R)$ is closed under arbitary intersection.

We now prove the following theorem.

THEOREM 7. *If R is a semi-prime ring, the annihilating ideals of $R[x]$ are precisely those ideals of the form $\mathfrak{a}[x]$, where $\mathfrak{a}$ is an annihilating ideal of R.*

This follows readily from the Corollary to Theorem 4 and the corresponding result for left ideals. If A is any ideal in $R[x]$ and $\mathfrak{a}$ is the ideal in R generated by all coefficients of elements of A, then $A^{rl} = \mathfrak{a}^{rl}[x]$. If, then, $A \in \mathfrak{A}(R[x])$ so that $A = A^{rl}$, it follows that $A = \mathfrak{a}^{rl}[x]$ and therefore $\mathfrak{a}^{rl} = \mathfrak{a}$. Hence $\mathfrak{a} \in A(R)$ and $A = \mathfrak{a}[x]$.

Conversely, if $\mathfrak{a} \in \mathfrak{A}(R)$ and we set $A = \mathfrak{a}[x]$, it follows that

$$A^{rl} = \mathfrak{a}^{rl}[x] = \mathfrak{a}[x] = A,$$

and $A \in \mathfrak{A}(R[x])$.

We may call $\mathfrak{A}(R)$ an *algebra* under the operations $\{\subseteq, \cap\}$. A minimal nonzero element of $\mathfrak{A}(R)$ will be called an *atom*. Since, by the preceding theorem, the algebras $\mathfrak{A}(R)$ and $\mathfrak{A}(R[x])$ are isomorphic, it follows that $\mathfrak{A}(R[x])$ has atoms if and only if $\mathfrak{A}(R)$ has atoms. We shall now make a few remarks about some relations between the structures of R and of $R[x]$ as determined by the isomorphism of $\mathfrak{A}(R)$ and $\mathfrak{A}(R[x])$. Throughout the rest of this section *we assume that the algebra $\mathfrak{A}(R)$ has atoms*. Our remarks are based on work of Johnson **(7)** whose structure theory for semi-prime rings involves a study of the prime right ideals and is more penetrating than any analysis depending only on the annihilating ideals.

If the atoms of $\mathfrak{A}(R)$ are $\mathfrak{b}_\alpha$, where α ranges over some index set, we shall call the ring sum of all these atoms the *support* of R and denote it by S. This is related to what Johnson calls the *base* of R, but we use a different term in order to avoid any confusion. It follows by the method of proof of Theorem 4.1 of **(7)** that $S = \bigcup_\alpha \mathfrak{b}_\alpha$, where $\bigcup\alpha$ indicates the finite direct sum of the

$\mathfrak{b}_\alpha$. Since the atoms of $R[x]$ are $\mathfrak{b}_\alpha[x]$, we know that $S[x]$ is then the support of the ring $R[x]$. Of course, this is the reason that the concept is of significance for our purposes.

It is known **(6)** that with a semi-prime ring R there is associated a unique semi-prime ring $N(R)$ such that (i) R is an ideal in $N(R)$, (ii) $R^r = 0$ in $N(R)$, and (iii) any ring T which contains R as an ideal and in which $R^r = 0$ is a subring of $N(R)$. Since $S = \cup_\alpha \mathfrak{b}_\alpha$, it follows **(7**, p. 387**)** that

$$N(S) = \sum_\alpha N(\mathfrak{b}_\alpha),$$

where $\sum_\alpha$ refers to the full direct sum.

Now $S \cap S^r = 0$ and therefore the sum of S and S^r is a direct sum $S \oplus S^r$. Moreover, $(S \oplus S^r)^r = 0$ and since $N(S \oplus S^r) = N(S) \oplus N(S^r)$, it follows that

$$S \oplus S^r \subseteq R \subseteq N(S) \oplus N(S^r). \tag{6}$$

We now investigate the question as to whether the relations (6) imply anything of the same sort about $R[x]$. From (6) it follows at once that

$$S[x] \oplus S^r[x] \subseteq R[x] \subseteq (N(S))[x] \oplus (N(S^r))[x]. \tag{7}$$

Moreover, we know that $S[x]$ is the support of $R[x]$, and the Corollary to Theorem 4 implies that $S^r[x] = (S[x])^r$. That is, the left side of (7) has the same relation to $R[x]$ that the left side of (6) has to R. Since S has no right annihilator in $N(S)$, S[x] has no right annihilator in $(N(S))[x]$, and hence $(N(S))[x] \subseteq N(S[x])$. Similarly, $(N(S^r))[x] \subseteq N(S^r[x])$, and thus

$$(N(S))[x] \oplus (N(S^r))[x] \subseteq N(S[x]) \oplus (N(S^r))[x]. \tag{8}$$

Moreover, $R[x]$ has no right annihilator in the left side of (8), so the second inclusion in (7) is apparently a little better than that obtained by simply applying relations (6) to the ring $R[x]$.

7. An example. We have established as Theorem 6 the "going up" and "going down" theorems for prime right ideals in semi-prime rings. The principal purpose of this section is to show by means of an example that there may be many prime right ideals in $R[x]$ not obtained from prime right ideals in R by the "going up" process.

The right ideal $\mathfrak{p}$ in a *prime* ring R is a prime right ideal if and only if $aRb \subseteq \mathfrak{p}$, where $a, b \in R$ with $b \neq 0$, implies that $a \in \mathfrak{p}$. Throughout this section R will be a *division* ring, and $R[x]$ is then a prime ring.

Let $c \in R$ and let us set $P = (x - c)R[x]$. It is not difficult to see that P is a semi-prime right ideal. However, if c is algebraic over the center of R, P is not a prime right ideal. For in this case there exists a nonzero polynomial $g(x)$ with coefficients in the center of R such that $g(c) = 0$, and $g(x)$ is therefore both right and left divisible by $x - c$. It follows that

$$1 \cdot R[x]g(x) = g(x)R[x] \subseteq P,$$

whereas $1 \notin P$.

From these remarks it follows that if R is a division ring and $R[x]$ has a prime right ideal of the form $(x - c)R[x]$, then c can not be algebraic over the center of R. We now give an example showing that there do exist prime right ideals of this form.

Henceforth, R will be the division ring of Hilbert **(4**, pp. 103ff**)**. For our purpose it is sufficient to know that R contains the field of rational numbers in its center and that it contains elements s and t such that $ts = 2st$. From this, it follows by induction that if α and β are nonnegative integers, then

$$s^{-\beta} t^{\alpha} s^{\beta} = 2^{\alpha\beta} t^{\alpha}. \tag{9}$$

We shall show that in $R[x]$ the right ideal $P = (x - t)R[x]$ is a prime right ideal. Accordingly, we assume that

$$g(x)R[x]h(x) \subseteq P \qquad h(x) \neq 0, \tag{10}$$

and shall show that $g(x) \in P$. By the division algorithm, we may write

$$g(x) = (x - t)q(x) + c,$$

where $c \in R$. From (10) it follows that

$$cR[x]h(x) \subseteq P. \tag{11}$$

The proof will be completed by showing that $c = 0$ and hence that $g(x) \in P$.

Let us assume that $c \neq 0$ and seek a contradiction. In particular, (11) then implies that $Rh(x) \subseteq P$. Thus, for every b in R, $bh(x)$ is left divisible by $x - t$. If $b \neq 0$ and $bh(x) = (x - t)F(x)$, it follows that

$$h(x) = b^{-1}(x - t)F(x) = (x - b^{-1}tb)b^{-1}F(x),$$

and $h(x)$ is left divisible by $x - b^{-1}tb$ for every nonzero $b \in R$. We proceed to show that this is impossible. Let

$$h(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n \qquad (a_0 \neq 0).$$

If $d \in R$, we may write

$$h_L(d) = d^n a_0 + d^{n-1}a_1 + \ldots + a_n,$$

and the factor theorem states that $h_L(d) = 0$ if and only if $h(x)$ is left divisible by $x - d$. Accordingly, we have that $h_L(b^{-1}tb) = 0$ for every nonzero b in R. In particular, $h_L(s^{-\beta}ts^{\beta}) = 0$ $(\beta = 0, 1, \ldots, n)$. Using (9), we then have the following system of equations:

$$\begin{aligned}
t^n a_0 \quad &+ t^{n-1}a_1 \quad &&+ \ldots + a_n = 0,\\
2^n t^n a_0 \quad &+ 2^{n-1}t^{n-1}a_1 \quad &&+ \ldots + a_n = 0,\\
2^{2n} t^n a_0 \quad &+ 2^{2(n-1)}t^{n-1}a_1 \quad &&+ \ldots + a_n = 0,\\
&\cdots \qquad\qquad &&\cdots\\
2^{n \cdot n} t^n a_0 \quad &+ 2^{n(n-1)}t^{n-1}a_1 \quad &&+ \ldots + a_n = 0.
\end{aligned}$$

However, since the determinant

$$\begin{vmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 \\ 2^n & 2^{n-1} & \cdot & \cdot & \cdot & 2 & 1 \\ 2^{2n} & 2^{2(n-1)} & & \cdot & \cdot & 2^2 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 2^{n \cdot n} & 2^{n(n-1)} & & \cdot & \cdot & 2^n & 1 \end{vmatrix}$$

is different from zero, it follows that

$$l^n a_0 = l^{n-1} a_1 = \ldots = a^n = 0,$$

and hence $a_i = 0$ $(i = 0, 1, \ldots, n)$. This implies that $h(x) = 0$ and we have the desired contradiction. It follows that $c = 0$ and hence that $g(x) \in P$, completing the proof that P is a prime right ideal in $R[x]$.

Since $R[x]$ is an integral domain and has a prime right ideal other than (0) and $R[x]$, it is known **(5)** that $R[x]$ does not have minimal prime right ideals. In the present example this also follows from the following statement whose proof we omit. If c_i $(i = 1, 2, \ldots, k)$ are distinct elements of the center of R, then

$$(x - c_1 t)(x - c_2 t) \ldots (x - c_k t) R[x]$$

is a prime right ideal in $R[x]$, and thus there exist prime right ideals which are generated by polynomials of arbitrarily high degree.

References

1. A. S. Amitsur, *The radical of a polynomial ring*, Can. J. Math., *8* (1956), 355–361.
2. R. Baer, *Radical ideals*, Amer. J. Math., *65* (1943), 537–568.
3. Jean-Claude Herz, *Sur les idéaux semi-premier ou parfaits. Étude des propriétés latticelles des idéaux semi-premiers*, C.R. Acad. Sci. Paris, *234* (1952), 1515–1517.
4. David Hilbert, *The Foundations of Geometry* (Chicago, 1910).
5. R. E. Johnson, *Prime rings*, Duke Math. J., *18* (1951), 799–809.
6. ———, *The imbedding of a ring as an ideal in another ring*, Duke Math. J., *20* (1953), 569–574.
7. ———, *Semi-prime rings*, Trans. Amer. Math. Soc., *76* (1954), 375–388.
8. J. Levitzki, *Prime ideals and the lower radical*, Amer. J. Math., *73* (1951), 25–29.
9. N. H. McCoy, *Prime ideals in general rings*, Amer. J. Math., *71* (1949), 823–833.
10. ———, *The prime radical of a polynomial ring*, Publ. Math. (Debrecen), *4* (1956), 161–162.
11. ———, *Annihilators in polynomial rings*, Amer. Math. Monthly, *64* (1957), 28–29.
12. M. Nagata, *On the theory of radicals in a ring*, Jour. Math. Soc. (Japan), *3* (1951), 330–344.
13. H. Tominaga, *Some remarks on radical ideals*, Math. Jour. Okayama Univ., *3* (1954), 139–142.

Smith College

ON BINOMIAL COEFFICIENT RESIDUES

J. B. ROBERTS

The number of binomial coefficients $\binom{u}{v}$, $0 \leqslant v \leqslant u < n$, which are congruent to j, $0 \leqslant j \leqslant p-1$, modulo the prime number p is denoted by $\theta_j(n)$. In this paper we give systems of simultaneous linear difference equations with constant coefficients whose solutions would yield the quantities $\theta_j(n)$ explicitly. In this direction we compute $\theta_j(n)$ in all cases for $p = 2$ and $\theta_j(p^k)$, $k > 0$, in all cases for $p = 3$ or 5. The complete explicit determination of $\theta_j(n)$ for arbitrary n is quite tedious for $p > 2$.

We also include various special results in the case $p = 2$ and prove that every prime divides "most" binomial coefficients in the sense that

$$\lim_{n \to \infty} \theta(n)/\theta_0(n) = 0$$

where

$$\theta(n) = \sum_{j=1}^{p-1} \theta_j(n).$$

1. Definitions. If c, a, s, k are constants satisfying $0 \leqslant a \leqslant c \leqslant p-1$, $1 \leqslant s \leqslant p^k$, $k > 0$, then the collection of all $\binom{u}{v}$ satisfying

$$cp^k \leqslant u < cp^k + s,\ ap^k \leqslant v \leqslant u + (a-c)p^k,$$

will be denoted by $(c, s, a)_k$. When we write $(c, s, a)_k$ we will assume that c, a, s, k satisfy the specified conditions unless stated explicitly to the contrary. For instance if we write $(0, s, a)_k$ this implies $0 \leqslant a \leqslant p-1$, $1 \leqslant s \leqslant p^k$, $k > 0$. Any collection $(c, s, a)_k$ will be called a *k-triangle*.

The k-triangle $(c, s, a)_k$ can be put into 1–1 correspondence with the k-triangle $(0, s, 0)_k$ by the mapping

$$\binom{u}{v} \leftrightarrow \binom{u - cp^k}{v - ap^k}.$$

Hence any two k-triangles can be put into 1 – 1 correspondence. Corresponding elements will be called *homologous*.

If K_1 and K_2 are two k-triangles and α is an integer such that $k_1 \equiv \alpha k_2 \pmod{p}$ whenever $k_1 \in K_1$ and $k_2 \in K_2$ are homologous we will write $K_1 \equiv \alpha K_2 \pmod{p}$.

Received October 2, 1956

2. A lemma of Lucas and applications. Our first lemma is a result of Lucas **(1**, p. 271**)**. A simple proof may be found in Glaisher **(2)**. We use p for a prime throughout.

LEMMA 1. *If in the scale of radix* p,

$$m = b_0 + b_1 p + \ldots + b_k p^k$$
$$n = a_0 + a_1 p + \ldots + a_k p^k,$$

then

$$\binom{n}{m} \equiv \binom{a_0}{b_0} \cdots \binom{a_k}{b_k} \pmod{p}.$$

(The quantity $\binom{r}{s} = 0$ when $s > r$.)

Before making use of this lemma we observe that by repeated use of the identity

$$\binom{a}{b} + \binom{a}{b+1} = \binom{a+1}{b+1}$$

and the almost obvious fact that

$$\binom{p^k}{m} \equiv 0 \pmod{p} \qquad (1 \leqslant m \leqslant p^k - 1,\ k > 0)$$

we are able to prove

LEMMA 2. *If* $n - p^k + 1 \leqslant m < p^k \leqslant n < 2p^k - 1$, $k > 0$, *then*

$$\binom{n}{m} \equiv 0 \pmod{p}.$$

We come now to our first application of Lemma 1.

LEMMA 3. *If* $0 \leqslant v \leqslant u$, $cp^k \leqslant u < (c+1)p^k$, $1 \leqslant c \leqslant p-1$ *and if* $\binom{u}{v}$ *is in none of the k-triangles* $(c, p^k, a)_k$, $0 \leqslant a \leqslant c$, *then*

$$\binom{u}{v} \equiv 0 \pmod{p}.$$

Proof. Since $\binom{u}{v}$ is not in $(c, p^k, a)_k$ for each a, $0 \leqslant a \leqslant c$, v must satisfy for some a, $0 \leqslant a \leqslant c-1$, the inequality

$$u + 1 + (a - c)p^k \leqslant v \leqslant (a+1)p^k - 1.$$

Since for each a, $0 \leqslant a \leqslant c-1$, this inequality is impossible when $u = (c+1)p^k - 1$ we can restrict attention to $u < (c+1)p^k - 1$. Now

$$u = a_0 + a_1 p + \ldots + a_{k-1} p^{k-1} + cp^k$$
$$v = b_0 + b_1 p + \ldots + b_{k-1} p^{k-1} + ap^k$$

and therefore by Lemma 1,

$$\binom{u}{v} \equiv \binom{a_0}{b_0} \cdots \binom{a_{k-1}}{b_{k-1}}\binom{c}{a} \equiv \binom{u-(c-1)p^k}{v-ap^k} \pmod{p}.$$

But since

$$\begin{aligned}(u-(c-1)p^k) - p^k + 1 = u + 1 - cp^k \leqslant v - ap^k \leqslant p^k - 1 \\ < p^k \leqslant u - (c-1)p^k < 2p^k - 1,\end{aligned}$$

we know by Lemma 3 that

$$\binom{u-(c-1)p^k}{v-ap^k} \equiv 0 \pmod{p}.$$

This completes the proof.

By this lemma we see that when $u \neq (c+1)p^k - 1$ there is always a v, $0 \leqslant v \leqslant u$, such that $\binom{u}{v}$ is divisible by p. It is interesting to note that for each u of the form $(c+1)p^k - 1$, $\binom{u}{v}$ is non-divisible by p for $0 \leqslant v \leqslant u$. Thus we state the

COROLLARY. *No* $\binom{u}{v}$, $0 \leqslant v \leqslant u$, *is divisible by* p *if and only if* u *is of the form* $(c+1)p^k - 1$ *where* $0 \leqslant c \leqslant p-1$.

Proof. The necessity is by the lemma. For the sufficiency we have

$$\begin{aligned}(c+1)p^k - 1 &= (p-1) + (p-1)p + \ldots + (p-1)p^{k-1} + cp^k \\ v &= b_0 + b_1 p + \ldots + b_{k-1}p^{k-1} + b_k p^k\end{aligned}$$

where $b_i \leqslant p-1$, $1 \leqslant i \leqslant k-1$ and $b_k \leqslant c$. Hence

$$\binom{(c+1)p^k - 1}{v} \equiv \binom{p-1}{b_0} \cdots \binom{p-1}{b_{k-1}}\binom{c}{b_k} \pmod{p}$$

by Lemma 1. But this right-hand side is not congruent to 0 modulo p. This completes the proof.

Another important application of Lemma 1 is the following

LEMMA 4.

$$(c, s, a)_k \equiv \binom{c}{a} (0, s, 0)_k \pmod{p}.$$

Proof. Let $\binom{u}{v}$ be in $(c, s, a)_k$. Then $cp^k \leqslant u < cp^k + s$, $ap^k \leqslant v \leqslant u + (a-c)p^k$ and we can write in radix p,

$$u = a_0 + a_1 p + \ldots + a_{k-1} p^{k-1} + cp^k$$
$$v = b_0 + b_1 p + \ldots + b_{k-1} p^{k-1} + ap^k.$$

Hence by Lemma 1,

$$\binom{u}{v} \equiv \binom{u - cp^k}{v - ap^k} \binom{c}{a} \pmod{p}.$$

Since $\binom{u-cp^k}{v-ap^k}$ runs over $(0, s, 0)_k$ as $\binom{u}{v}$ runs over $(c, s, a)_k$ the proof is complete.

COROLLARY. *The number of numbers in* $(c, s, a)_k$ *which are congruent to* $j \pmod{p}$, $1 \leqslant j \leqslant p - 1$, *is*

$$\theta_{j_a}(s)$$

where j_a *is that number satisfying*

$$1 \leqslant j_a \leqslant p - 1, j_a \binom{c}{a} \equiv j \pmod{p}.$$

Proof. By the lemma a number in $(c, s, a)_k$ is congruent to j modulo p if and only if $\binom{c}{a}$ times its homologous element in $(0, s, 0)_k$ is congruent to j modulo p. Since

$$j_a \binom{c}{a} \equiv j \pmod{p}$$

the number of possibilities is the number of j_a in $(0, s, 0)_k$ and this is just

$$\theta_{j_a}(s).$$

3. The main recursion relation. Utilizing Lemma 3 we see that for $0 \leqslant c \leqslant p - 1$, $1 \leqslant s \leqslant p^k$ all of those $\binom{u}{v}$, $0 \leqslant v \leqslant u$, $cp^k \leqslant u < cp^k + s$, which are not congruent to zero modulo p are in one of the $c + 1$ k-triangles $(c, s, a)_k$, $0 \leqslant a \leqslant c$. Therefore $\theta_j(cp^k + s) - \theta_j(cp^k)$ is just the number of elements congruent to j modulo p contained in these k-triangles. By Lemma 4 this number is

$$\sum_{a=0}^{c} \theta_{j_a}(s).$$

Defining $e_{qj}(c)$, $1 \leqslant q \leqslant p - 1$, to be the number of j_a, $0 \leqslant a \leqslant c$, which are equal to q, the above sum becomes

$$\sum_{q=1}^{p-1} e_{qj}(c)\, \theta_q(s).$$

But $e_{qj}(c)$ is just the number of solutions of the congruence

$$\binom{c}{x} q \equiv j \pmod{p},$$

which number is, by definition,

$$\theta_{j\bar{q}}(c+1) - \theta_{j\bar{q}}(c)$$

where $\bar{q}$ is the reciprocal of q modulo p. Hence we have the following theorem setting forth our main recursion relation.

THEOREM 1. *If* $0 \leqslant c \leqslant p-1$, $1 \leqslant s \leqslant p^k$, $k > 0$, $q\bar{q} \equiv 1 \pmod{p}$ *then*

$$\theta_j(cp^k + s) = \theta_j(cp^k) + \sum_{q=1}^{p-1} (\theta_{j\bar{q}}(c+1) - \theta_{j\bar{q}}(c))\, \theta_q(s).$$

Remembering the definition of $\theta(n)$ we have under the hypotheses of the theorem the following

COROLLARY 1. $\theta(cp^k + s) = \theta(cp^k) + (c+1)\theta(s)$.

Proof. For each q, $1 \leqslant q \leqslant p-1$, the residues modulo p of the numbers $\bar{q}, 2\bar{q}, \ldots, (p-1)\bar{q}$ are the numbers $1, 2, \ldots, p-1$ in some order. Using this fact and the theorem we obtain

$$\begin{aligned}
\theta(cp^k + s) &= \sum_{j=1}^{p-1} \theta_j(cp^k + s) \\
&= \sum_{j=1}^{p-1} \theta_j(cp^k) + \sum_{q=1}^{p-1} \sum_{j=1}^{p-1} (\theta_{j\bar{q}}(c+1) - \theta_{j\bar{q}}(a))\, \theta_q(s) \\
&= \theta(cp^k) + (\theta(c+1) - \theta(c))\, \theta(s).
\end{aligned}$$

Since $\theta(c+1) - \theta(c) = c+1$, because c is smaller than p, the proof is complete.

COROLLARY 2. *If* $0 \leqslant c \leqslant p$, $k > 0$ *then*

(a) $$\theta_j(cp^k) = \sum_{q=1}^{p-1} \theta_{j\bar{q}}(c)\, \theta_q(p^k);$$

(b) $$\theta(cp^k) = \tfrac{1}{2} c(c+1)\, \theta(p^k).$$

Proof. (a) This is true for $c = 0$ or $k = 0$ so we suppose $c > 0$, $k > 0$. Now taking $s = p^k$ the theorem gives, for $1 \leqslant c \leqslant p$,

$$\begin{aligned}
\theta_j(cp^k) &= \sum_{i=1}^{c} (\theta_j(ip^k) - \theta_j((i-1)p^k)) \\
&= \sum_{i=1}^{c} \sum_{q=1}^{p-1} (\theta_{j\bar{q}}(i) - \theta_{j\bar{q}}(i-1))\, \theta_q(p^k) \\
&= \sum_{q=1}^{p-1} \sum_{i=1}^{c} (\theta_{j\bar{q}}(i) - \theta_{j\bar{q}}(i-1))\, \theta_q(p^k) \\
&= \sum_{q=1}^{p-1} \theta_{j\bar{q}}(c)\, \theta_q(p^k).
\end{aligned}$$

(b)
$$\theta(cp^k) = \sum_{j=1}^{p-1} \theta_j(cp^k)$$
$$= \sum_{q=1}^{p-1} \left(\sum_{j=1}^{p-1} \theta_{j\bar{q}}(c) \right) \theta_q(p^k)$$
$$= \theta(c)\,\theta(p^k) = \tfrac{1}{2}\,c(c+1)\,\theta(p^k).$$

COROLLARY 3. (a) *If* $k > 0$, $1 \leqslant j \leqslant p-1$ *then*

$$\theta_j(p^k) = \sum_{q=1}^{p-1} \theta_{j\bar{q}}(p)\,\theta_q(p^{k-1})\ ;$$

(b) *If* $k \geqslant 0$ *then*

$$\theta(p^k) = (\tfrac{1}{2}p(p+1))^k.$$

Proof. (a) Taking $c = p$ in Cor. 2 (a) gives

$$\theta_j(p^{k+1}) = \sum_{q=1}^{p-1} \theta_{j\bar{q}}(p)\,\theta_q(p^k),\ k > 0$$

and this is equivalent with (a).

(b) This is obvious for $k = 0$. If true up to some $k > 0$ then by Cor. 2(b),

$$\theta(p^{k+1}) = \tfrac{1}{2}\,p(p+1)\,\theta(p^k) = (\tfrac{1}{2}\,p(p+1))^{k+1}.$$

This completes the proof.

By repeated application of these corollaries we are able to give an explicit expression for $\theta(n)$. This we do in the next corollary.

COROLLARY 4. *If* $n = a_0 + a_1 p + \ldots + a_k p^k$, $0 \leqslant a_i \leqslant p-1$ *then*

$$\theta(n) = \tfrac{1}{2} \sum_{i=0}^{k} a_i((a_i+1)\ldots(a_k+1))\,(\tfrac{1}{2}\,p(p+1))^i.$$

Theorem 1 and its corollaries determine the $\theta_j(n)$, $1 \leqslant j \leqslant n$, as solutions of a system of linear difference equations with constant coefficients. The quantity

$$\theta_0(n) = \tfrac{1}{2}n(n+1) - \theta(n).$$

In general the calculations needed to compute explicitly the $\theta_j(n)$ are prohibitive. However we perform some calculations in this direction in the next section.

4. $\theta_j(p^k)$ for $p = 3, 5$. The simplest case to deal with is $p = 2$. In this case we can compute $\theta_j(n)$ for arbitrary n. The details will be given in the next section where some other aspects of our results for $p = 2$ are discussed.

When $p = 3$, since $(j\bar{q})q \equiv j \pmod 3$, we have

$$1\,\bar{1} \equiv 2\bar{2} \equiv 1,\ 1\bar{2} \equiv 2\bar{1} \equiv 2 \pmod 3.$$

By direct examination we find $\theta_1(3) = 5$, $\theta_2(3) = 1$. We now obtain from Cor. 3(a) of Theorem 1 the following pair of simultaneous difference equations

$$\theta_1(3^k) - 5\,\theta_1(3^{k-1}) - \theta_2(3^{k-1}) = 0,$$
$$\theta_2(3^k) - \theta_1(3^{k-1}) - 5\theta_2(3^{k-1}) = 0.$$

Solving these equations using the empirical initial conditions

$$\theta_1(1) = \theta_2(3) = 1, \theta_1(3) = 5, \theta_2(1) = 0$$

we obtain

$$\theta_1(3^k) = \tfrac{1}{2}\,(6^k + 4^k), \theta_2(3^k) = \tfrac{1}{2}(6^k - 4^k).$$

From these it follows that

$$\theta_0(3^k) = \tfrac{1}{2}\,3^k(3^k + 1) - 6^k.$$

In a similar way with $p = 5$ we find a system of four linear difference equations in four unknowns. Using the suitable initial conditions we obtain:

$$\theta_1(5^k) = \tfrac{1}{4}(15^k + 9^k + (8 - i)^k + (8 + i)^k),$$
$$\theta_2(5^k) = \tfrac{1}{4}(15^k - 9^k - i(8 - i)^k + i(8 + i)^k),$$
$$\theta_3(5^k) = \tfrac{1}{4}(15^k - 9^k + i(8 - i)^k - i(8 + i)^k),$$
$$\theta_4(5^k) = \tfrac{1}{4}(15^k + 9^k - (8 - i)^k - (8 + i)^k).$$

From these it follows that

$$\theta_0(5^k) = \tfrac{1}{2}\,5^k(5^k + 1) - 15^k.$$

5. The case $p = 2$. In the case $p = 2$, Cor. 4 of Theorem 1 reads as follows:
(1) If $n = 2^{\alpha_1} + \ldots + 2^{\alpha_r}$, $\alpha_1 > \ldots > \alpha_r$, then

$$\theta(n) = \sum_{i=1}^{r} 2^{i-1} \cdot 3^{\alpha_i}.$$

Since every n is of one of the three forms:

(i) $2^{\alpha_1} + \ldots + 2^{\alpha_r}$ with $\alpha_1 > \ldots > \alpha_r > 0$;
(ii) $2^{\alpha_1} + \ldots + 2^{\alpha_r} + 2^s + 2^{s-1} + \ldots + 2 + 1$ with $\alpha_1 > \ldots > \alpha_r > s+1$;
(iii) $2^s + 2^{s-1} + \ldots + 2 + 1$

we can use (1) to compute $\theta(n + 1) - \theta(n)$ finding its values in the three cases to be 2^r, 2^{s+r}, 2^{s+1} respectively. Hence we have the result:

(2) the number of odd $\binom{n}{m}$ for fixed n and $0 \leqslant m \leqslant n$ is equal to 2^s where s is the number of non-zero digits in the binary expansion of n.

This result was proved by Glaisher **(2)** from our Lemma 1. From this we have the special result, which can be proved in a very nice way directly **(3**, p. 15 problem 12 and the solution pp. 97-98**)**, that the nth row of Pascal's triangle consists of odd numbers exclusively if and only if n is a power of 2. This special case is also an immediate consequence of the corollary to Lemma 3.

If we let $\theta_n = \theta_1(n+1) - \theta_1(n)$ and $E_n = \theta_0(n+1) - \theta_0(n)$ we have the result:

(3) $E_n < \theta_n$ if and only if $n + 1 < 2^{1+s}$ where s is the number of non-zero digits in the binary expansion of n. In all other cases $E_n > \theta_n$.

The first statement in (3) follows from (2) since $E_n - \theta_n = n + 1 - 2^s$. In order to prove the second part of (3) suppose the contrary. That is, suppose $E_n = \theta_n$ for some n. Then by (2), $n + 1 = 2^{1+s}$ or $n = 2^s + \ldots + 1$. But then the number of non-zero digits in the binary expansion of n is $s + 1$. This is a contradiction and therefore $E_n \neq \theta_n$ for all n.

We include one other result whose proof we omit.

(4) $\theta_1(n) > \theta_0(n)$ if and only if $1 < n < 18$.

6. "Most binomial coefficients are divisible by a given prime". In this section we prove the

THEOREM 2.

$$\lim_{n\to\infty} \theta(n)/\theta_0(n) = 0.$$

Proof. Clearly $\theta(n)$ and $\theta_0(n)$ are non-decreasing functions of n. Hence if $p^k \leqslant n < p^{k+1}$ then, using Cor. 3(b) of Theorem 1,

$$\theta(n)/\theta_0(n) \leqslant \theta(p^{k+1})/\theta_0(p^k)$$

$$= \binom{p+1}{2}^{k+1}\left\{\binom{p^k+1}{2} - \binom{p+1}{2}^k\right\}^{-1} = p(p+1)\left\{\left(\frac{2p}{p+1}\right)^k + \left(\frac{2}{p+1}\right)^k - 2\right\}^{-1},$$

and this tends to 0 as $n \to \infty$.

REFERENCES

1. L. E. Dickson, *History of the Theory of Numbers*, vol. 1 (Chelsea, New York, 1952).
2. J. W. L. Glaisher, *On the residue of a binomial theorem coeffcient with respect to a prime modulus*, Quart. J. Math., *30* (1899), 150–156.
3. I. M. Vinogradoff, *An Introduction to the Theory of Numbers*, (London and New York, Pergamon Press, 1955).

Wesleyan University,
Middletown, Conn.

and

Reed College
Portland, Oregon.

COMBINATORIAL PROPERTIES OF MATRICES OF ZEROS AND ONES

H. J. RYSER

1. Introduction. This paper is concerned with a matrix A of m rows and n columns, all of whose entries are 0's and 1's. Let the sum of row i of A be denoted by r_i $(i = 1, \ldots, m)$ and let the sum of column i of A be denoted by s_i $(i = 1, \ldots, n)$. It is clear that if τ denotes the total number of 1's in A

$$\tau = \sum_{i=1}^{m} r_i = \sum_{i=1}^{n} s_i.$$

With the matrix A we associate the *row sum vector*

$$R = (r_1, \ldots, r_m),$$

where the ith component gives the sum of row i of A. Similarly, the *column sum vector* S is denoted by

$$S = (s_1, \ldots, s_n).$$

We begin by determining simple arithmetic conditions for the construction of a (0, 1)-matrix A having a given row sum vector R and a given column sum vector S. This requires the concept of majorization, introduced by Muirhead. Then we apply to the elements of A an elementary operation called an interchange, which preserves the row sum vector R and column sum vector S, and prove that any two (0, 1)-matrices with the same R and S are transformable into each other by a finite sequence of such interchanges. The results may be rephrased in the terminology of finite graphs or in the purely combinatorial terms of set and element. Applications to Latin rectangles and to systems of distinct representatives are studied.

2. Maximal matrices and majorization. Let

$$\delta_i = (1, \ldots, 1, 0, \ldots, 0)$$

be a vector of n components with 1's in the first r_i positions, and 0's elsewhere. A matrix of the form

$$\bar{A} = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_m \end{bmatrix}$$

is called *maximal*, and we refer to $\bar{A}$ as the *maximal form* of A. The maximal $\bar{A}$ may be obtained from A by a rearrangement of the 1's in the rows of A. Also by inverse row rearrangements one may construct the given A from $\bar{A}$.

Received July 1, 1956. This work was sponsored in part by the Office of Ordnance Research.

Let $\bar{R} = (\bar{r}_1, \ldots, \bar{r}_m)$ and $\bar{S} = (\bar{s}_1, \ldots, \bar{s}_n)$ be the row sum and column sum vectors of $\bar{A}$. Evidently

$$R = \bar{R}.$$

Moreover, it is clear that the row sum vector R uniquely determines $\bar{A}$, and hence $\bar{S}$. Indeed, $\tau = \sum r_i = \sum \bar{s}_i$ constitute conjugate partitions of τ.

Consider two vectors $S = (s_1, \ldots, s_n)$ and $S^* = (s_1^*, \ldots, s_n^*)$, where the s_i and s_i^* are nonnegative integers. The vector S is *majorized* by S^*,

$$S \prec S^*,$$

provided that with the subscripts renumbered **(5; 3)**:

(1) $\quad s_1 \geqslant \ldots \geqslant s_n,\ s_1^* \geqslant \ldots \geqslant s_n^*$;

(2) $\quad s_1 + \ldots + s_i \leqslant s_1^* + \ldots + s_i^*, \qquad i = 1, \ldots, n-1;$

(3) $\quad s_1 + \ldots + s_n = s_1^* + \ldots + s_n^*.$

For the vectors S and $\bar{S}$ associated with the matrices A and $\bar{A}$, respectively, we prove that

$$S \prec \bar{S}.$$

We renumber the subscripts of the s_i of A so that

$$s_1 \geqslant s_2 \geqslant \ldots \geqslant s_n.$$

For $\bar{A}$, we already have

$$\bar{s}_1 \geqslant \bar{s}_2 \geqslant \ldots \geqslant \bar{s}_n.$$

Now A must be formed from $\bar{A}$ by a shifting of 1's in the rows of $\bar{A}$. But for each $i = 1, \ldots, n-1$, the total number of 1's in the first i columns of $\bar{A}$ cannot be increased by a shifting of 1's in the rows of $\bar{A}$. Hence

$$s_1 + \ldots + s_i \leqslant \bar{s}_1 + \ldots + \bar{s}_i,$$

$i = 1, \ldots, n-1$. Moreover,

$$s_1 + \ldots + s_n = \bar{s}_1 + \ldots + \bar{s}_n,$$

whence we conclude that $S \prec \bar{S}$.

THEOREM 2.1[1]. *Let the matrix $\bar{A}$ be maximal and have column sum vector $\bar{S}$. Let S be majorized by $\bar{S}$. Then by rearranging 1's in the rows of $\bar{A}$, one may construct a matrix A having column sum vector S.*

Without loss of generality, we may assume that the column sums of A satisfy $s_1 \geqslant s_2 \geqslant \ldots \geqslant s_n$. We construct the desired A inductively by columns by a rearrangement of the 1's in the rows of $\bar{A}$.

[1] *Added in proof.* The author has been informed recently that Theorem 2.1 was obtained independently by Professor David Gale. His investigations concerning this theorem and certain generalizations are to appear in the Pacific Journal of Mathematics.

By hypothesis, $S < \bar{S}$, whence $s_1 \leqslant \bar{s}_1$. If $s_1 = \bar{s}_1$, we leave the first column of $\bar{A}$ unchanged. Suppose that $s_1 < \bar{s}_1$. We may rearrange 1's in the rows of $\bar{A}$ to obtain s_1 1's in the first column, unless

$$\bar{s}_2 > s_1, \ldots, \bar{s}_n > s_1.$$

But if these inequalities hold, then

$$\bar{s}_1 + \ldots + \bar{s}_n > ns_1 \geqslant s_1 + \ldots + s_n = \bar{s}_1 + \ldots + \bar{s}_n,$$

which is a contradiction.

Let us suppose then that the first t columns of A have been constructed, and let us proceed to the construction of column $t + 1$. We have then given an m by n matrix

$$[\eta_1, \ldots, \eta_t, \quad \eta_{t+1}, \ldots, \eta_n],$$

where the number of 1's in column η_i is s_i $(i = 1, \ldots, t)$. Let the number of 1's in column η_j be s'_j $(j = t + 1, \ldots, n)$. We may suppose that

$$s'_{t+1} \geqslant s'_{t+2} \geqslant \ldots \geqslant s'_n.$$

Two cases arise.

Case 1. $$s_{t+1} < s'_{t+1}.$$

In this case, remove 1's from column η_{t+1} by row rearrangements, and place the 1's in columns $\eta_{t+2}, \ldots, \eta_n$. If sufficiently many 1's may be removed from η_{t+1} in this manner, then we are finished. Suppose then that there remain e 1's in column $t + 1$, with

$$s_{t+1} < e \leqslant s'_{t+1},$$

and that no further 1's may be removed by this procedure. Then there must exist an integer $w > 0$ such that

$$s_{t+1} + \ldots + s_n = (n - t)e + w.$$

But

$$\begin{aligned} &s_{t+1} < e, \\ &s_{t+2} \leqslant s_{t+1} < e, \\ &\cdot \\ &\cdot \\ &\cdot \\ &s_n < e. \end{aligned}$$

Therefore

$$(n - t)e + w = s_{t+1} + \ldots + s_n < (n - t)e,$$

which is a contradiction.

Case 2. $$s'_{t+1} < s_{t+1}.$$

By row rearrangements, insert 1's into column η_{t+1} from columns $\eta_{t+2}, \ldots, \eta_n$. If sufficiently many 1's may be inserted in this manner, then we are finished. Suppose then that there remain e 1's in column $t+1$ with

$$s'_{t+1} \leqslant e < s_{t+1},$$

and that no further 1's may be inserted by our procedure.

Let the matrix at this stage of the construction process be denoted by

$$[e_{rs}].$$

If now

$$e_{r,t+1} = 0,$$

then

$$e_{rj} = 0, \qquad (j = t+1, \ldots, n).$$

Suppose that some

$$e_{rj} = 1, \qquad j > t+2.$$

Then either

$$e_{rk} = 1, \qquad (k = 1, \ldots, t+1),$$

or else for some k, $1 \leqslant k \leqslant t$,

$$e_{rk} = 0.$$

Consider the case in which $e_{rk} = 0$. Since $s_k \geqslant s_{t+1} > e$, there must exist

$$e_{pk} = 1, \quad e_{p,t+1} = 0.$$

Interchanging $e_{rj} = 1$ and $e_{rk} = 0$, and interchanging $e_{pk} = 1$ and $e_{p,t+1} = 0$, we see that $s_1, \ldots, s_t$ are left unaltered, and that e is increased by 1.

Continue to increase e by transformations of this variety. Suppose that all such transformations have been applied and that e still satisfies

$$s'_{t+1} \leqslant e < s_{t+1}.$$

But now it is no longer possible to move a 1 from columns $t+2, \ldots, n$ into columns $1, 2, \ldots, t+1$. This means that

$$s_1 + \ldots + s_t + e = \bar{s}_1 + \ldots + \bar{s}_t + \bar{s}_{t+1}.$$

But then

$$s_1 + \ldots + s_{t+1} \leqslant \bar{s}_1 + \ldots + \bar{s}_{t+1} = s_1 + \ldots + s_t + e,$$

whence $s_{t+1} \leqslant e$, which is a contradiction. This completes the proof.

The preceding theorem has a variety of applications. For example, let the $(0, 1)$-matrix A of m rows and n columns contain exactly $\tau = km$ 1's, where k is a positive integer. Let the column sum vector of A be $S = (s_1, \ldots, s_n)$. Then there exists an m by n matrix A^* composed of 0's and 1's with exactly k 1's in each row, and column sum vector S. For let $\bar{A}$ be m by n, with all 1's in the first k columns and 0's elsewhere. If $\bar{S}$ denotes the column sum vector of $\bar{A}$, then $S \prec \bar{S}$, and the desired A^* may be constructed from $\bar{A}$.

In this connection we mention the following result arising in the study of the completion of Latin rectangles **(1; 7)**. Let A be a (0, 1)-matrix of r rows and n columns, $1 \leqslant r < n$. Let there be k 1's in each row of A, and let the column sums of A satisfy $k - (n - r) \leqslant s_i \leqslant k$. Then $n - r$ rows of 0's and 1's may be adjoined to A to obtain a square matrix with exactly k 1's in each row and column **(7)**. To prove this it suffices to construct an $n - r$ by n matrix A^* of 0's and 1's with exactly k 1's in each row, and column sum vector $(k - s_1, \ldots, k - s_n)$. By the remarks of the preceding paragraph, such a construction is always possible.

3. Interchanges. We return now to the m by n matrix A composed of 0's and 1's, with row sum vector R and column sum vector S. We are concerned with the 2 by 2 submatrices of A of the types

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

An *interchange* is a transformation of the elements of A that changes a specified minor of type A_1 into type A_2, or else a minor of type A_2 into type A_1, and leaves all other elements of A unaltered. Suppose that we apply to A a finite number of interchanges. Then by the nature of the interchange operation, the resulting matrix A^* has row sum vector R and column sum vector S.

THEOREM 3.1. *Let A and A^* be two m by n matrices composed of* 0's *and* 1's, *possessing equal row sum vectors and equal column sum vectors. Then A is transformable into A^* by a finite number of interchanges.*

The proof is by induction on m. For $m = 1$ and 2, the theorem is trivial. The induction hypothesis asserts the validity of the theorem for two (0, 1)-matrices of size $m - 1$ by n.

We attempt to transform the first row of A into the first row of A^* by interchanges. If we are successful, the theorem follows at once from the induction hypothesis. Suppose that we are not successful and that we denote the transformed matrix by A'. For notational convenience, we simultaneously permute the columns of A' and A^* and designate the first row of A' by

$$(\delta_r, \eta_s, \delta_t, \eta_t)$$

and the first row of A^* by

$$(\delta_r, \eta_s, \eta_t, \delta_t).$$

Here δ_r and δ_t are vectors of all 1's with r and t components, respectively, and η_s and η_t are 0 vectors with s and t components, respectively. Thus we have been successful in obtaining agreement between the two rows in the positions labelled δ_r and η_s, but have been unable to obtain agreement in the positions labelled δ_t and η_t. We may suppose, moreover, that these 2t positions of disagreement are the minimal number of disagreements obtainable among

all attempts to transform the first row of A into the first row of A^* by interchanges.

Let A'_{m-1} and A^*_{m-1} denote the matrices composed of the last $m-1$ rows of A' and A^*, respectively. The row sum vectors of A'_{m-1} and A^*_{m-1} are equal. Also corresponding columns of A'_{m-1} and A^*_{m-1} below the positions labelled δ_r and η_s have equal sums. Let α_i denote the $(r+s+i)$th column of A'_{m-1}, and let β_i denote the $(r+s+t+i)$th column of A'_{m-1}, where $i = 1, \ldots, t$. Let $\alpha_1^*, \ldots, \alpha_t^*$ and $\beta_1^*, \ldots, \beta_t^*$ denote the corresponding columns of A^*_{m-1}. Let a_i, b_i, a_i^*, b_i^* denote the column sums of $\alpha_i, \beta_i, \alpha_i^*, \beta_i^*$, respectively.

Now in A'_{m-1} we cannot have simultaneously a 0 in the position determined by row j and column α_i and a 1 in the position determined by row j and column β_i. For if this were the case, we could perform an interchange and reduce the $2t$ disagreements in the first row of A'. Hence $a_i \geqslant b_i$. Moreover, $a_i^* = a_i + 1$ and $b_i^* = b_i - 1$, whence

$$a_i^* - b_i^* = a_i - b_i + 2 \geqslant 2.$$

In A^*_{m-1}, consider columns α_i^* and β_i^*. There exists a row of A^*_{m-1} that has a 1 in column α_i^* and a 0 in column β_i^*. Replace the 1 by 0 and the 0 by 1, and let such a replacement be made for each $i = 1, \ldots, t$. We obtain in this way a new matrix $\bar{A}_{m-1}$ whose row and column sum vectors are equal to those of A'_{m-1}. By the induction hypothesis, we may transform A'_{m-1} into $\bar{A}_{m-1}$ by interchanges. However, these interchanges applied to A' will allow us to perform further interchanges and make the first rows of the transformed A' and A^* coincide. Hence the theorem follows.

Let $\mathfrak{A}$ denote the class of all $(0, 1)$-matrices of m rows and n columns, with row sum vector R and column sum vector S. The *term rank* ρ of A in $\mathfrak{A}$ is the order of the greatest minor of A with a nonzero term in its determinant expansion **(6)**. This integer is also equal to the minimal number of rows and columns that contain collectively all of the nonzero elements of A **(4)**. A $(0, 1)$-matrix $A = [a_{rs}]$ may be considered an incidence matrix distributing n elements $x_1, \ldots, x_n$ into m sets $S_1, \ldots, S_m$. Here $a_{ij} = 1$ or 0 according as x_j is or is not in S_i. From this point of view the term rank of a matrix is a generalization of the concept of a system of distinct representatives for subsets $S_1, \ldots, S_m$ of a finite set N **(2)**. Indeed, the subsets $S_1, \ldots, S_m$ possess a system of distinct representatives if and only if $\rho = m$.

THEOREM 3.2. *Let $\bar{\rho}$ be the minimal and $\bar{\bar{\rho}}$ the maximal term rank for the matrices in $\mathfrak{A}$. Then there exists a matrix in $\mathfrak{A}$ possessing term rank ρ, where ρ is an arbitrary integer on the range*

$$\bar{\rho} \leqslant \rho \leqslant \bar{\bar{\rho}}.$$

For an interchange applied to a matrix in $\mathfrak{A}$ either changes the term rank by 1 or else leaves it unaltered. But by Theorem 3.1, we may transform the matrix of term rank $\bar{\bar{\rho}}$ into the matrix of term rank $\bar{\rho}$. This implies that there exists a matrix in $\mathfrak{A}$ of term rank ρ.

References

1. Marshall Hall, *An existence theorem for Latin squares*, Bull. Amer. Math. Soc., *51* (1945), 387–388.
2. P. Hall, *On representatives of subsets*, J. Lond. Math. Soc., *10* (1935), 26–30.
3. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge, 1952).
4. Dénes König, *Theorie der endlichen und unendlichen Graphen* (New York, 1950).
5. R. F. Muirhead, *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters*, Proc. Edinburgh Math. Soc., *21* (1903), 144–157.
6. Oystein Ore, *Graphs and matching theorems*, Duke Math. J., *22* (1955), 625–639.
7. H. J. Ryser, *A combinatorial theorem with an application to Latin rectangles*, Proc. Amer. Math. Soc., *2* (1951), 550–552.

Ohio State University

A CLASS OF NON-DESARGUESIAN PROJECTIVE PLANES

D. R. HUGHES

1. Introduction. In **(7)**, Veblen and Wedderburn gave an example of a non-Desarguesian projective plane of order 9; we shall show that this plane is self-dual and can be characterized by a collineation group of order 78, somewhat like the planes associated with difference sets. Furthermore, the technique used in **(7)** will be generalized and we will construct a new non-Desarguesian plane of order p^{2n} for every positive integer n and every odd prime p. To do this, we need a result due to Zassenhaus **(8)** that there exists a near-field which is not a field of every order p^{2n}, p an odd prime, whose center is a field of order p^n. However there are some significant differences between the case $p^{2n} = 9$ and all other cases, and these lead to some unsolved problems.

Furthermore, we shall show that none of the planes constructed in this fashion can be coordinatized by Veblen-Wedderburn systems (with either distributive law) and that each such plane possesses a (non-linear) planar ternary ring whose additive loop is an elementary abelian group.

2. Construction of planes. A finite *left Veblen-Wedderburn system* (left V-W system) is a finite set R containing at least the two distinct elements 0 (zero) and 1 (one), together with two binary operations, addition $(+)$ and multiplication $(\cdot)$ (where we often write ab for $a \cdot b$), all satisfying:

(1) R is a group under addition, with "identity" 0;

(2) The non-zero elements of R form a loop under multiplication, with identity 1;

(3) $0x = x0 = 0$, all $x \in R$;

(4) The left distributive law is valid in R:

$$a(b + c) = ab + ac, \qquad a, b, c \in R.$$

Similarly, a right V-W system satisfies the right distributive law, in place of (4). A left (right) V-W system with associative multiplication is a *left (right) near-field*. Throughout this paper we shall omit the term "finite," as all V-W systems considered will be finite; it has been shown that the addition of any V-W system (indeed, even the infinite ones) is commutative, and (for the finite case only) is elementary abelian. Hence R always has order equal to a power of a prime. (See **(4)** for proofs.)

Received October 20, 1955; in revised form March 20, 1956. This research was carried out while the author was a National Science Foundation Fellow. Presented to the American Mathematical Society, February 25, 1956.

Any (left or right) V-W system is a linear planar ternary ring **(2; 3)**. If R is a (left or right) V-W system which is not a field then the plane π coordinatized by R is non-Desarguesian and contains a distinguished line or point (depending on which distributive law is present) which is moved by no collineation; a pair of anti-isomorphic V-W systems coordinatize planes which are duals. In **(8)** Zassenhaus has determined all (finite) near-fields, as well as their automorphism groups, and in particular, has shown that for any odd prime p and any positive integer n there is a near-field R (which is not a field) of order p^{2n} whose center is a field of order p^n. (By "center" we mean here the set of all elements $z \in R$ such that $zx = xz$ for all $x \in R$.)

Let R be a left near-field (which is not a field) of order p^{2n} containing a field F of order p^n as its center; let $q = p^{2n} + p^n + 1$. For the sake of simplicity in the argument, we introduce the set V of all ordered triples (x, y, z) where x, y, z are in R, and let V_0 be the subset of V consisting of all triples all of whose entries are in F. Then V is a left vector space over R (or over F), and V_0 is a left vector space over F. Suppose A is a (non-singular) linear transformation of V, as a vector space over R, with the following additional properties:

(i) $V_0A = V_0$,

(ii) if v is in V, then $vA^q = kv$ for some k in R, $k \neq 0$,

(iii) if v_0 is in V_0, $v_0 \neq (0, 0, 0)$, and if $v_0A^m = kv_0$ where k is in F, $k \neq 0$, then $m \equiv 0 \pmod{q}$.

Then, using (i), if x, y, z are in R, we can write:

$$(1)\quad (x, y, z)A^m = (a_{11}x + a_{12}y + a_{13}z, a_{21}x + a_{22}y + a_{23}z, a_{31}x + a_{32}y + a_{33}z),$$

where the a_{ij} are in F (and the a_{ij} depend only on m, of course). Furthermore, if A has the form given by (1), as a linear transformation only of V_0 over F, and if A satisfies (ii) and (iii) (as they apply to V_0 and F), then certainly A is a linear transformation of V over R and satisfies (i), (ii), and (iii). So the existence of A depends only on the existence of a linear transformation A_0 of V_0 over F satisfying:

(iv) if v_0 is in V_0, $v_0 \neq (0, 0, 0)$, then $v_0A_0^m = kv_0$, where k is in F, $k \neq 0$, if and only if $m \equiv 0 \pmod{q}$.

We shall return to the question of the existence of A_0 after some discussion of the use of A.

Given R, F, and A as above, let us construct a set π of points and lines, with an incidence relation (i.e., point on line or line contains point, etc.), as follows. The points of π will be the elements of V, excepting the element $(0, 0, 0)$, with the identification $(x, y, z) = (kx, ky, kz)$, for any non-zero k in R. The lines of π are the formal symbols L_tA^m, where either $t = 1$ or t is in R, t not in F, and where the only identifications will be $L_tA^k = L_tA^m$ if $k \equiv m \pmod{q}$. Incidence is as follows: $v = (x, y, z)$ is on L_t if $x + yt + z = 0$, while L_tA^m contains just those points vA^m such that v is on L_t. By simple counting, it is seen that π contains $p^{4n} + p^{2n} + 1$ points and the same number

of lines, and that each line is incident with $p^{2n} + 1$ distinct points (we have not yet shown, of course, that an arbitrary pair of distinct lines differ at all as sets of points). Thus, from **(5)**, if we can show that each two distinct lines have exactly one point in common, then π will be a projective plane of order p^{2n}.

Now the existence of A_0 is assured by the work of Singer **(6)**, since the cyclic collineation given by him is easily seen to be nothing but a linear transformation of the vector space V_0, and to satisfy (iv). Indeed, Singer states as much, since he represents his projective plane by "homogeneous" coordinates from the field F (i.e., ordered triples with the identification we have used above), and then, without explicitly making use of the vector space, shows the existence of a linear transformation A_0 whose qth power (and no smaller positive power) maps an element (x, y, z) onto an element (kx, ky, kz), $k \neq 0$.

Thus if we succeed in showing that the set π constructed above is a projective plane, it will even contain a subplane π_0 of order p^n, consisting exactly of those points (x, y, z) for which x, y, z are in F, and of the lines L_1A^m. (Thus π_0 will be Desarguesian.) Furthermore, whether π is a projective plane or not, each mapping A^m is a collineation of π: if the point P is on the line L, then PA^m is on LA^m. From this last remark, to show that any pair of distinct lines intersect in exactly one point, it will be sufficient to show that any pair of distinct lines L_tA^m and L_s intersect in exactly one point. In what follows, we assume that equation (1) takes the following form for A^{-m}:

$$(2)\quad (x, y, z)A^{-m} = (a_{11}x + a_{12}y + a_{13}z, a_{21}x + a_{22}y + a_{23}z, a_{31}x + a_{32}y + a_{33}z).$$

Let us consider the intersection of the distinct lines L_tA^m and L_s; if (x, y, z) is on both lines, then $(x, y, z)A^{-m}$ is on L_t, so we have:

$$(3)\quad (a_{11}x + a_{12}y + a_{13}z) + (a_{21}x + a_{22}y + a_{23}z)t + (a_{31}x + a_{32}y + a_{33}z) = 0,$$
$$(4)\quad x + ys + z = 0,$$

where each a_{ij} is in F. Solving equation (4) for x and substituting in (3), we have:

$$(5)\qquad yu + za + (yv + zb)t = 0,$$

where:

$$(6)\qquad u = a_{12} + a_{32} - (a_{11} + a_{31})s, \quad v = a_{22} - a_{21}s,$$
$$a = a_{13} + a_{33} - (a_{11} + a_{31}), \quad b = a_{23} - a_{21}.$$

Note that a, b are in F. We now have several cases.

Case 1; $b \neq 0$. Then equation (5) can be written :

$$(yv + zb)b^{-1}a + y(u - vb^{-1}a) + (yv + zb)t = 0,$$

utilizing the fact that a, b^{-1} are in the center. This becomes:

$$(7)\qquad (yv + zb)\,(b^{-1}a + t) + y(u - vb^{-1}a) = 0.$$

It is easy to see that if $t = 1$ then (5) and **(4)** have a unique common solution for the point (x, y, z). So we assume $t \neq 1$, and thus in (7), t is not in F, and hence $w = b^{-1}a + t \neq 0$. So (7) becomes $(yv + zb)w = -y(u - vb^{-1}a)$, or, multiplying through by w^{-1} and collecting terms,

$$y[v + (u - vb^{-1}a)w^{-1}] + zb = 0. \tag{8}$$

Since not both the coefficients in (8) can be zero (i.e., $b \neq 0$), (8) and (4) have a unique common solution for the point (x, y, z).

Case II; $b = 0$, $a \neq 0$. Then (5) becomes:

$$y(u + vt) + za = 0. \tag{9}$$

But since $a \neq 0$, (9) and (4) have also a unique common point.

Case III; $a = b = 0$. That is to say:

$$a_{13} + a_{33} = a_{11} + a_{31}, \qquad a_{23} = a_{21}. \tag{10}$$

But now consider the element $v_0 = (1, 0, -1)$; from (10) it is immediate that $v_0A^{-m} = cv_0$, where $c = a_{11} - a_{13}$. Necessarily $c \neq 0$, since A^{-m} is not singular. Thus $m \equiv 0 \pmod{q}$, so $L_tA^m = L_t$, and then it is easy to see that the distinct lines L_t and L_s have only the point $(1, 0, -1)$ in common.

We have thus completed the proof that π is a projective plane of order p^{2n}, and we note that π possesses a cyclic collineation group (generated by A) of order $q = p^{2n} + p^n + 1$. Furthermore, the collineation A fixes no point or line of π, so if we succeed in showing that π is non-Desarguesian, then it even has the stronger property that it cannot be coordinatized by a V-W system, in any manner whatsoever. For use in § 4, we note that the line L_tA^m of π can be represented by an equation:

$$xa + yb + zc + (xa' + yb' + zc')t = 0, \tag{11}$$

where a, b, c, a', b', c', are all in F.

3. Collineation groups. As pointed out above, π possesses a cyclic group of collineations of order $q = p^{2n} + p^n + 1$. Furthermore, if θ is any automorphism of the near-field R, and if θ fixes every element in F, then the mapping $(x, y, z) \to (x\theta, y\theta, z\theta)$, $L_tA^m \to L_{t\theta}A^m$, is a collineation of π. The near-field of order 9 possesses the non-abelian group of order 6 as a group of automorphisms **(8)** and this group necessarily fixes every element in the subfield of order three. Thus the plane has a group of collineations of order 78, which is the direct product of the cyclic group of order 13 and the non-abelian group of order 6. In **(7)** this collineation group is given explicitly, and it is pointed out that the group is transitive and regular on the points and lines of π which are not in the subplane π_0 of order three (π_0 is mapped into itself by all of the 78 collineations). We shall abstract from this situation before analysing the plane of order 9 in more detail.

Let π be a projective plane containing a proper subplane π_0 of finite order m. Suppose G is a group of collineations of π such that π_0 is mapped into itself by every element of G, and such that G is transitive and regular on the points and lines of π that are not in π_0. Since any line of π_0 contains a point P not in π_0, every point of π that is not in π_0 is on exactly one line of π_0; for any point of π, not in π_0, is an image of P under some element of G. Let π have order n; as every point of π is on at least one of the $m^2 + m + 1$ lines of π_0, n is finite, for no finite set of lines carries all of the points of an infinite plane. Each line of π_0 carries $n - m$ points not in π_0. So there are

$$(n - m)(m^2 + m + 1) + m^2 + m + 1 = n^2 + n + 1$$

points in π. Solving this for n, and noting that $n \neq m$, it is elementary that $n = m^2$. If L_1 and L_2 are lines of π_0, containing the points P_1 and P_2, respectively, where $P_1, P_2 \notin \pi_0$, then since $P_1x = P_2$ for some $x \in G$, we have $L_1x = L_2$ since L_2 is the only line of π_0 containing P_2. Thus G is transitive on the lines, and similarly on the points, of π_0. The order of G is equal to the number of points of π which are not in π_0; i.e., G has order $m^4 + m^2 + 1 - (m^2 + m + 1) = m^4 - m$. Let the points and lines of π that are not in π_0 be called tangent points and tangent lines. Let P_0 be some fixed tangent point and let J_0 be some fixed tangent line; we can assume that P_0 is on J_0. Let K_0 be the unique line of π_0 which contains P_0, and let Q_0 be the unique point of π_0 which is on J_0. Let D be the subset of G consisting of all x such that P_0x is on J_0, let E consist of all x such that P_0x is on K_0, let F consist of all x such that J_0x contains Q_0, and let D_0 consist of all x such that Q_0x is on K_0.

We observe that E and F are subgroups of G. For if e is in E, then P_0e is a point of K_0, and must lie on exactly one line of π_0; hence this line must be K_0, so $K_0e = K_0$, and E is the subgroup of G which fixes K_0. Similarly, F is the subgroup which fixes Q_0. If K is any line of π_0, then K is fixed by some conjugate (in G) of E, and if Q is any point of π_0, then Q is fixed by some conjugate of F. Furthermore, $FD_0E = D_0$; E and F have order $m^2 - m$, D contains m^2 elements and D_0 contains $(m + 1)(m^2 - m)$ elements.

THEOREM 1. (i) *If a is not in E, then $a = d_1d_2^{-1}$ for a unique d_1, d_2 in D; if a is in E, $a \neq 1$, then $a \neq d_1d_2^{-1}$ for any d_1, d_2 in D.*

(ii) *The left cosets of F can all be represented as dF for a unique d in D, or as $d_0^{-1}F$ for d_0 in D_0, but not both.*

(iii) *If a is not in F, then $a = d_1d_2^{-1}$ for d_1, d_2 in D_0, where d_1, d_2 are uniquely determined up to a common right multiple by an element of E (i.e., d_1e, d_2e, where e is in E, are also in D_0 and $a = (d_1e)(d_2e)^{-1}$, but a only has representations of this form).*

(iv) $FD_0E = D_0$.

Proof. (i) If a is not in E, then $P_0 \cdot P_0 a$ is a line $J_0 b$ for a unique b. Hence $b^{-1} = d_2$ in D, $ab^{-1} = d_1$ in D, so $a = d_1 d_2^{-1}$; since b is unique, it is easy to see that d_1, d_2 are unique. Conversely, if a is in E, $a \neq 1$, then the impossibility of $a = d_1 d_2^{-1}$ is easy to demonstrate.

(ii) As in (i), consideration of the point $R = J_0 \cdot J_0 a$, where a is not in F, shows that $a = d_1^{-1} d_2$ for a unique pair d_1, d_2 in D, and if a is in F, $a \neq 1$, then $a \neq d_1^{-1} d_2$ for any d_1, d_2 in D. Hence if d_1, d_2 are in D and $d_1 F = d_2 F$, then $d_1^{-1} d_2$ is in F, so $d_1 = d_2$. Thus the m^2 cosets dF, for d in D, are all distinct.

Consider the points P_0 and $Q_0 a$; the line $L = P_0 \cdot Q_0 a$ is a line $J_0 b$ if and only if b^{-1} is in D and $Q_0 a = Q_0 b$; i.e., if and only if a^{-1} is in $b^{-1} F$, where b^{-1} is in D. On the other hand, if $L = K_0 b$, then b is in E and ab^{-1} is in D_0, so a is in $D_0 b \subseteq D_0 E = D_0 = F D_0$. So a^{-1} is in $d_0^{-1} F$, where d_0 is in D_0. This proves (ii).

(iii) Consider the line $K_0 b = Q_0 \cdot Q_0 a$, where a is not in F. We have, as in (i), $a = d_1 d_2^{-1}$, where d_1, d_2 are in D_0, but d_1, d_2 are not unique: since $K_0 b = K_0 eb$ for any e in E, it is easy to see that $d_1 e$, $d_2 e$ (both in D_0) also represent a as $a = (d_1 e)(d_2 e)^{-1}$, and that all such representations are of this form.

(iv) It has already been pointed out that (iv) is satisfied.

Now, without giving the proofs (which are straightforward but time-consuming), we remark that the existence of a group G of order $m^4 - m$, containing subgroups E and F of order $m^2 - m$ and two subsets D and D_0, with m^2 and $(m + 1)(m^2 - m)$ elements respectively, all satisfying (i)-(iv) of Theorem 1, implies the existence of the projective plane π. Points are designated by (a) for all a in G, and (Fa) for all cosets Fa of F; lines are $[Db]$ for all b in G, and $[Eb]$ for all cosets Eb of E. Incidence is given by the rules: (a) is on $[Db]$ if a is in Db; (a) is on $[Eb]$ if a is in Eb; (Fa) is on $[Db]$ if b is in Fa; (Fa) is on $[Eb]$ if a is in $D_0 b$.

Suppose such a group possesses an automorphism λ with the properties $E\lambda = E$, $F\lambda = F$, $d^{-1}\lambda$ is in D if d is in D, $d_0^{-1}\lambda$ is in D_0 if d_0 is in D_0. Then the mapping T defined below is a duality of the plane π, the simple proof of which statement we omit.

$$T: \quad (a) \to [D \cdot a\lambda] \qquad [Db] \to (b\lambda)$$
$$(Fa) \to [E \cdot a\lambda] \qquad [Eb] \to (F \cdot b\lambda).$$

In the case of the plane of order 9 given by Veblen and Wedderburn, G can be taken as the group of order 78 mentioned above, and $E = F$ is the subgroup of order 6; note that E is normal in G. The subsets D and D_0 depend upon our choice of P_0 and J_0, so we reproduce the plane below. The points are the symbols $A_i, B_i, \ldots, G_i$, $i = 0, 1, \ldots, 12$, and we give seven of the lines; the remaining lines are found by successively adding one to the subscripts, reducing modulo 13 (indeed, this operation corresponds to the collineation A):

$$
\begin{array}{lllllllllll}
L_1 & : A_0 & A_1 & A_3 & A_9 & B_0 & C_0 & D_0 & E_0 & F_0 & G_0 \\
L_j & : A_0 & B_1 & B_8 & D_3 & D_{11} & E_2 & E_8 & E_6 & G_7 & G_9 \\
L_{2j} & : A_0 & C_1 & C_8 & E_7 & E_9 & F_3 & F_{11} & G_2 & G_8 & G_6 \\
L_{1+j} & : A_0 & B_7 & B_9 & D_1 & D_8 & F_2 & F_4 & F_6 & G_3 & G_{11} \\
L_{2+2j} & : A_0 & B_2 & B_4 & B_6 & C_3 & C_{11} & E_1 & E_8 & F_7 & F_9 \\
L_{1+2j} & : A_0 & C_7 & C_9 & D_2 & D_4 & D_6 & E_3 & E_{11} & F_1 & F_8 \\
L_{2+j} & : A_0 & B_3 & B_{11} & C_2 & C_4 & C_6 & D_7 & D_9 & G_1 & G_8 .
\end{array}
$$

Then the group $E = F$ corresponds to the permutations:

$$E = \{1, (BDG)(CEF), (BGD)(CFE), (BC)(DF)(EG), (BE)(CD)(FG), (BF)(CG)(DE)\}.$$

Let us choose B_2 as P_0 and L_{2+2j} as J_0. Then K_0 is L_1A^2, Q_0 is A_0, and:

$$D = \{1, A^3, A^4, (BC)(DF)(EG)A, (BC)(DF)(EG)A^9, (BE)(CD)(FG)A^6, (BE)(CD)(FG)A^{12}, (BF)(CG)(DE)A^5, (BF)(CG)(DE)A^7\},$$

$$D_0 = \{A^2, A^6, A^{10}, A^{11}\} \cdot E.$$

Every element of G can be represented in the form eA^k, where e is in E. Let λ be the automorphism $eA^k \to eA^{-k}$; then $E\lambda = E$, $d^{-1}\lambda$ is in D if d is in D and $d_0^{-1}\lambda$ is in D_0 if d_0 is in D_0. So (see above) the plane is self-dual.

This plane has some further interesting properties, some of which we will mention. Each one of the following sets of three points is on a line, and it is easy to see that the whole set of seven points, together with the seven lines joining them, forms a subplane of order two:

$$A_3A_9B_0;\ A_2A_3D_2;\ A_2A_9G_6;\ B_0D_2G_6;\ A_3D_{11}G_6;\ A_2B_0D_{11};\ A_9D_2D_{11}.$$

Since no element of G fixes all of the seven points, this means that there are at least 78 distinct subplanes of order two and every point of the plane is in at least four subplanes of order two.

The duality T given above has 22 absolute points (i.e., points on their image line): $A_1, A_8, A_9, A_{12}, B_2, B_4, B_{10}, C_2, C_4, C_{10}, \ldots, G_2, G_4, G_{10}$. The line L_1A^5 contains two absolute points A_1, A_8, while L_1A contains four absolute points A_1, B_2, D_4, G_{10}. Thus the duality is not "regular," giving a negative answer to a question raised by Baer **(1)**.

Finally, π contains "ovals"; i.e., sets of ten points, no three of which are on a line. An example of one is the set of points:

$$A_0, A_1, A_6, A_7, B_2, C_8, C_9, D_8, D_9, E_2.$$

The oval given above even has the strong property that there is another duality of the plane (distinct from T) whose absolute points are exactly the points of the oval (the duality can, in fact, be constructed from the oval). Hence the collineation group of the plane consists of more than the 78 collineations given above; for if T_1 and T_2 are dualities of any projective plane, then T_1T_2 is a collineation of the plane. The author has not been able to determine the full collineation group of the plane given above.

4. Coordinate rings. We shall use the planar ternary ring coordinatizing a projective plane to derive some more properties of the class of planes constructed in § 2; in particular, we shall show that these planes are all non-Desarguesian, and cannot be coordinatized by Veblen-Wedderburn systems. The particular formulation of the ternary ring will follow the lines of **(3)**, but we briefly review the basic idea, which differs a bit from Hall's technique in **(2)**. In **(4)** will also be found a good deal of material on these ternary rings. R is a nonempty set containing at least the two distinct elements 0 and 1, and points are symbols (x, y), (m), (∞), where $m, x, y \in R$, and ∞ is a symbol not in R; lines are symbols $[m, k]$, $[\infty, (k, 0)]$, L_∞, where $m, k \in R$. A ternary operation F is defined so that (x, y) is on $[m, k]$ if and only if $F(m, x, y) = k$; the other rules of incidence are: (x, y) is on $[\infty, (x, 0)]$, (m) is on $[m, k]$ and L_∞, (∞) is on $[\infty, (k, 0)]$ and L_∞. Then the ternary function F satisfies certain axioms, which will be found in **(2, 3; 4)**. Addition is defined by $a + b = F(1, a, b)$ and multiplication by $a \cdot b = ab = F(a, b, 0)$; the ring is called *linear* if $F(a, b, c) = ab + c$ for all $a, b, c \in R$. It is well known that every planar ternary ring for a Desarguesian plane is an associative division ring, and, in particular, is linear.

Throughout the rest of this section let π be a projective plane of order p^{2n}, constructed from the left near-field R, as in § 2. We shall coordinatize π so as to construct one of its planar ternary rings, and in what follows, the ordered triples have the same meaning as in § 2. Let (∞) be $(0, 0, 1)$; (0) be $(1, 0, 0)$; $(0, 0)$ be $(0, 1, 0)$; and let (1) be $(1, 0, -1)$. The x-axis, y-axis, and L_∞ are then all lines of the form L_1A^k, and, in particular, they are respectively $z = 0$, $x = 0$, and $y = 0$. The points on the y-axis are (in the old representation) all of the form $(0, 1, v)$, so let $(0, 1, v)$ be $(0, v)$ in our new coordinate system. Every line through (1), or $(1, 0, -1)$, is a line of the form $x + yt + z = 0$. The point $(v, 0)$ on the x-axis will be the point $(u, 1, 0)$ which is collinear with $(1, 0, -1)$ and $(0, 1, v)$; but $(1, 0, -1)$ and $(0, 1, v)$ are on the line L_{-v} if $v \notin F$, whence $u + 1(-v) + 0 = 0$, or $u = v$. If $v \in F$, then it is immediate that $u = v$. So $(v, 0)$ is $(v, 1, 0)$.

The point (m) will be the point on L_∞ which is collinear with $(1, 1, 0)$ and $(0, 1, m)$, and will be a point $(1, 0, v)$. Let $xa + yb + zc + (xa' + yb' + c'z)t = 0$ be the line joining $(1, 1, 0)$ and $(0, 1, m)$. Then:

$$a + b + (a' + b')t = 0 \tag{1}$$

$$b + mc + (b' + mc')t = 0. \tag{2}$$

Since $a, a', b, b' \in F$, (1) implies that either $t = 1$ and $a + a' = -(b + b')$, or $t \neq 1$ and $a + b = a' + b' = 0$. If $t = 1$, then $a + a' + v(c + c') = 0$, and (2) becomes

$$a + a' + (-m)(c + c') = 0;$$

thus $v = -m$. If $t \neq 1$, then $a + vc + (a' + vc')t = 0$, and (2) becomes $-a + mc + (-a' + mc')t = 0$, so again $v = -m$. Thus (m) is the point $(1, 0, -m)$.

Consider the point (u, v) which lies on the line J_1 joining $(0, 0, 1)$ and $(u, 1, 0)$ and on the line J_2 joining $(1, 0, 0)$ and $(0, 1, v)$. Then:

$$J_1 : xa + yb + zc + (xa' + yb' + zc')t = 0, \tag{3}$$
$$J_2 : xd + ye + zf + (xd' + ye' + zf')s = 0,$$

where

$$c + c't = d + d's = ua + b + (ua' + b')t = e + vf + (e' + vf')s = 0. \tag{4}$$

Thus as before, $t = 1$ and $c = -c'$, or $t \neq 1$ and $c = c' = 0$; similarly, $s = 1$ and $d = -d'$, or $s \neq 1$ and $d = d' = 0$.

There are four cases to check, but all of them are easy. If $t \neq 1$, $s \neq 1$, then it is a simple verification, using (4), that $(u, 1, v)$ is on both of the lines J_1 and J_2. If $t = 1$, $s \neq 1$, then J_1 is

$$x(a + a') + y(b + b') = 0,$$

where $u(a + a') + b + b' = 0$, so $(u, 1, v)$ is on J_1. The line J_2 becomes

$$ye + zf + (ye' + zf')s = 0,$$

where $e + vf + (e' + vf')s = 0$, so $(u, 1, v)$ is on J_2. In all cases, we find that $(u, 1, v)$ is on both J_1 and J_2, so (u, v) is $(u, 1, v)$.

So we have:

THEOREM 2. *If π is coordinatized as above, then (u, v) is $(u, 1, v)$, (m) is $(1, 0, -m)$ and (∞) is $(0, 0, 1)$.*

Now we shall investigate the ternary ring for π, where we use $T(a, b, c)$ for the ternary operation, and let $a \oplus b = T(1, a, b)$, $a \circ b = T(a, b, 0)$. In order to find the value of $T(m, u, v)$, we consider the line L which contains $(1, 0, -m)$ and $(u, 1, v)$, and let $(0, 1, k)$ be the intersection of L with the y-axis; then $k = T(m, u, v)$. Let L be the line $xa + yb + zc + (xa' + yb' + zc')t = 0$. Then:

$$a - mc + (a' - mc')t = 0 \tag{5}$$
$$ua + b + vc + (ua' + b' + vc')t = 0. \tag{6}$$

THEOREM 3. *For all a and b, $a \oplus b = a + b$.*

Proof. Let $m = 1$ in (5). Then $a - c + (a' - c')t = 0$. So if $t = 1$, we have $a + a' = c + c'$, and (6) becomes $u(a + a') + (b + b') + v(c + c') = 0$, or $(u + v)(a + a') + (b + b') = 0$. But then the point $(0, 1, u + v)$ is on L, so $k = u \oplus v = u + v$. If $t \neq 1$, then $a = c$ and $a' = c'$, and (6) becomes

$$(u + v)\, a + b + [(u + v)a' + b']t = 0,$$

whence again $(0, 1, u + v)$ is on L, so $u \oplus v = u + v$.

THEOREM 4. *The ternary ring (R, T) is not linear.*

Proof. Referring to (5) and (6), let m and u be arbitrary, $u \neq 0$, and let v be chosen so that $k = T(m, u, v) = 0$. Then $(0, 1, 0)$ is on L, so $b + b't = 0$.

Suppose $t = 1$, whence $b = -b'$. Then (5) is $a + a' - m(c + c') = 0$, and (6) is

$$u(a + a') + v(c + c') = 0,$$

and this implies $u^{-1}v = -m$, or $um + v = 0$. Suppose $t \neq 1$, whence $b = b' = 0$. Then (5) is $a - mc + (a' - mc')t = 0$, and (6) can be written

$$a + u^{-1}vc + (a' + u^{-1}vc')t = 0,$$

and so again $u^{-1}v = -m$, or $um + v = 0$.

Now assume that (R,T) is linear. For arbitrary m and u, $u \neq 0$, let $p = m \circ u$. Then $m \circ u \oplus (-p) = T(m, u, -p) = 0$, so by the above, $um + (-p) = 0$, or $um = p = m \circ u$. Thus (R,T) is anti-isomorphic to the near-field R, and so (R,T) is itself a near-field. As pointed out earlier, a (finite) near-field plane possessing a collineation moving every point and line is necessarily Desarguesian. But π does possess such a collineation; it is obvious that A fixes no point or line of π. Hence (R,T) is a field, so R is also a field, and this is contradictory; thus (R,T) cannot be linear.

COROLLARY. *The plane π is non-Desarguesian.*

Proof. See the proof of Theorem 4. Or note that every ternary ring for a Desarguesian plane is linear, whence by Theorem 4, π is not Desarguesian.

Besides the class of projective planes given in this paper, all finite planes known at the present time are coordinatizable by ternary rings which are (at least) V-W systems. So, in a sense, the class of planes given here are the "weakest" finite planes known. Aside from having prime-power order, these planes share a much stronger property with the V-W system planes however: they can be coordinatized by a planar ternary ring whose addition forms an elementary abelian group, or equivalently, there is a complete set of mutually orthogonal latin squares associated with the plane which contains among its squares the Cayley table of an elementary abelian group.

Other questions about the class of planes, which are answered in § 3 for the plane of order 9, include: are the planes self-dual and does there exist a collineation group G with the properties discussed in § 3? The automorphism groups of the near-fields used in this paper, with the exception of the one of order 9, are always too small to yield enough collineations to construct the group G. The author has checked a plane of the class, of order 25, and it does not possess such a group G of collineations. Finally, what other finite planes possess such a collineation group G?

References

1. Reinhold Baer, *Polarities in finite projective planes*, Bull. Amer. Math. Soc. *52* (1946), 77–93.
2. Marshall Hall, Jr., *Projective planes*, Trans. Amer. Math. Soc. *54* (1943), 229–277.
3. D. R. Hughes, *Planar division neo-rings*, Trans. Amer. Math. Soc. *80* (1955), 502–527.
4. Günter Pickert, *Projektive Ebenen* (Berlin 1955).
5. H. J. Ryser, *A note on a combinatorial problem*, Proc. Amer. Math. Soc. *1* (1950), 422–424.
6. James Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc. *43* (1938), 377–385.
7. O. Veblen and J. H. M. Wedderburn, *Non-desarguesian and non-pascalian geometries*, Trans. Amer. Math. Soc. *8* (1907), 379–388.
8. H. Zassenhaus, *Ueber endliche Fastkörper*, Abh. Math. Sem. Univ. Hamburg, *11* (1935), 187–220.

The Ohio State University

TRANSITIVITIES IN PROJECTIVE PLANES

T. G. OSTROM

Introduction. A perspectivity in a projective plane is a collineation which leaves some line pointwise fixed. Baer **(4)** has considered the coordinatization of planes which admit certain groups of perspectivities. André **(1;2;3)** has made an extensive study of the Veblen-Wedderburn plane in terms of its perspectivities. The author has shown **(7)** that finite doubly transitive planes are Desarguesian if the number of points on a line is $n + 1$ where n is an odd non-square.

In Part 1, we consider certain products of perspectivities. Our main results are certain uniqueness theorems for the group of permutations of points on a line induced by perspectivities.

In Part 2, we consider the coordinatization of planes admitting two $p-L$ transitive groups. (See definition in Part 1.) In various cases, depending upon the kinds of perspectivities and the relative location of the centers and axes, we obtain planes which can be coordinatized by various types of algebraic systems.

In Part 3, we consider the exceptional cases not taken care of in the author's paper on double transitivity. We dispose of the case where n is an odd power of 2. For all values of n, we show that transitivity on quadrangles implies that every quadrangle generates a Desarguesian subplane.

In Part 4, we show that every finite plane contains Desarguesian configurations.

1. Uniqueness theorems on perspectivities.

Definition. A collineation which leaves fixed every line through some point p (called the *center*) and every point on some line L (called the *axis*) is a *perspectivity*.

We shall make use of the well-known fact that the above definition is redundant in the sense that if every line through p is fixed then there is some line L such that every point on L is fixed **(4)**.

Definition. A perspectivity in which the center is on the axis will be called an *elation*. A perspectivity in which the center is not on the axis will be called a *homology*.

Definition. Let Σ be the group of perspectivities with center p and axis L. Then, if Σ is transitive on the non-fixed points of each line through p, the plane is said to be $p-L$ *transitive*.

Received September 10, 1956.

LEMMA 1. *Let ρ and σ be two homologies with the same center p, and let the axes of ρ and σ intersect in some point $q \neq p$. Let r be any point on the line pq other than p or q. Then if r is fixed by $\rho\sigma$, $\rho\sigma$ is a perspectivity with center p and axis qr.*

Proof. Since $\rho\sigma$ leaves every line through p fixed, $\rho\sigma$ is a perspectivity with center p. Moreover, both q and r are fixed by $\rho\sigma$. For a perspectivity all fixed points other than the center must lie on the axis. Hence qr is the axis of $\rho\sigma$.

THEOREM 1. *Let ρ and σ be two homologies with the same center p, and let the axes of ρ and σ intersect in some point $q \neq p$. If some point r (other than p or q) on the line pq has the same image r_1 under both ρ and σ, then ρ and σ produce the same permutation of points on the line rq.*

Proof. By Lemma 1, $\rho\sigma^{-1}$ leaves every point of rq fixed.

COROLLARY. *Suppose that the plane is $p-L$ transitive, where $p \notin L$. Let $L_1 \neq L$ be some line through p intersecting L in some point $q \neq p$. Let ρ be a collineation fixing both p and q. Let Σ be the group of $p-L$ perspectivities, and let G be the induced permutation group on the line pq. Then $\rho^{-1}\Sigma\rho$ induces an automorphism of G.*

Proof. If r and r_1 are two points of the line pq, other than p or q, there is exactly one member of Σ (and hence of G) which carries r into r_1 **(4)**. But $\rho^{-1}\Sigma\rho$ is another transitive group of perspectivities with center p and some axis through q. It follows from Theorem 1 that the group G_1 induced by $\rho^{-1}\Sigma\rho$ is the same as G.

LEMMA 2. *Let ρ and σ be two elations with the same center p. Let r be any point $\neq p$. Then if r is fixed by $\rho\sigma$, $\rho\sigma$ is an elation with center p and axis rp.*

Proof. Since every line through p is fixed by $\rho\sigma$, $\rho\sigma$ is a perspectivity with center p. Let L be the axis of ρ. If the axis of $\rho\sigma$ does not go through p, it contains some point $q \neq p$ on L. The point q is a fixed point of ρ and $\rho\sigma$, and hence is a fixed point of σ. But the only fixed points of an elation are those on the axis. Thus $pq = L$ is the axis of σ. We conclude that the axis of $\rho\sigma$ does go through p and $\rho\sigma$ is an elation. The lemma follows from the fact that the only fixed points of an elation are those on the axis.

THEOREM 2. *Let ρ and σ be two elations with the same center p. If $r \to r_1$ under both ρ and σ, then ρ and σ produce the same permutation of points on pr.*

COROLLARY. *Suppose that the plane is $p-L$ transitive, where $p \in L$. Let $L_1 \neq L$ be some line through p. Let ρ be a collineation fixing p and L_1. Let Σ be the group of $p-L$ perspectivities, and let G be the induced permutation group on L_1. Then $\rho^{-1}\Sigma\rho$ is an automorphism of G.*

THEOREM 3. *Let ρ and σ be two elations with the same center p but with different axes L and M. Then ρ and σ are of the same period.*

Proof. Consider the elation $\rho\sigma$. Points on L are permuted in cycles whose length is equal to the period of σ and points on M are permuted in cycles whose length is equal to the period of ρ. But the lengths of both of these cycles are equal to the period of $\rho\sigma$. (We are here using the fact that the non-fixed points of a perspectivity are all permuted in cycles of the same length. This in turn depends upon the fact that a perspectivity can have no fixed points off the axis, other than the center.)

2. Perspectivities and coordinate systems. In this section, we shall be using Hall's method **(5)** of coordinatizing a plane, in which the equation of a line is represented by a ternary operation $y = x \cdot m \circ b$. "Addition" and "multiplication," respectively, are defined in terms of the ternary by $xm = x \cdot m \circ 0$ and $x + b = x \cdot 1 \circ b$. It is not, in general, true that $x \cdot m \circ b = xm + b$. If this identity holds, the coordinatization is said to be linear.

We shall depart from our convention of using small letters for points and capital letters for lines to the extent that we shall use "A" to indicate the center of the pencil x = constant, and "B" for the center of the pencil y = constant.

Our purpose in this part is to consider the coordinate systems which arise (with the proper choice of the quadrangle of reference) in various cases in which the plane admits two p–L transitive groups. We shall refer to some of Baer's results on p–L transitivity **(4)**. Baer's method of coordinatization is different from that of Hall; however, they amount to the same thing if $a \cdot 1 = a$ in Baer's system or $a \cdot (-1) = -a$ in Hall's system. By and large, the same methods apply in both cases, so we shall apply some of Baer's results to Hall's system without further explanation.

For convenience, we include the definition of the more common coordinatizing algebras, and some theorems already known.

Definition. An algebraic system consisting of two binary operations, addition and multiplication, is called a *Veblen-Wedderburn system*, or *quasifield*, if it satisfies the following conditions:

1. Addition is an Abelian group.
2. $(b + c)a = ba + ca$.
3. $a \cdot 0 = 0 \cdot a = 0$.
4. $1 \cdot a = a \cdot 1 = a$.
5. $ay = b$ has a unique solution y if $a \neq 0$.
6. $-xb + xa = c$ has a unique solution x if $a \neq b$.

The system is called a *nearfield* if

7. Multiplication is associative.

If condition 2 is replaced by the other distributive law, $a(b + c) = ab + ac$, with corresponding interchange of right and left multiplication in 5 and 6,

we shall call the system a *dual Veblen-Wedderburn system* or *dual nearfield* as is appropriate.

Definition. A plane which can be coordinatized by a Veblen-Wedderburn system with $x \cdot m \circ b = xm + b$ is a *Veblen-Wedderburn plane.*

It is well known that the dual of a Veblen-Wedderburn plane can be coordinatized by a dual Veblen-Wedderburn system.

Definition. A pair of points p and q on the line at infinity of a Veblen-Wedderburn plane is said to be *admissible* if the plane is p–L transitive for some line L through q, where $L \neq L_\infty$.

THEOREM 4 (André **(3)**). *There are exactly the following possibilities for the set Z of all admissible pairs of points of a Veblen-Wedderburn plane* π:

I. *Z is empty.*

II. *Z consists of exactly one pair* (p,p). *In this case,* π *is a plane over a distributive quasifield, which is not a left alternative field.*

III. *Z consists of all pairs* (p,p) *on* L_∞. *In this case,* π *is a plane over a left alternative field which is not a skewfield.*

IV. *Z consists of exactly two pairs* (p, q) *and* (q, p) *with* $q \neq p$. *In this case,* π *is a plane over a near field which contains more than* 9 *elements and is not a skewfield.*

V. *For each* p *on* L_∞ *there exists one* $q \neq p$ *such that* (p, q) *and* (q, p) *belong to Z. In this case,* π *is the plane over the nearfield of order* 9.

VI. *Z consists of all pairs of points on* L_∞. *In this case,* π *is Desarguesian.*

THEOREM 5 (Baer **(4)**). *If the plane is A-*L_∞ *transitive, then* (i) *addition is a group and* (ii) *the coordinatization is linear.*

THEOREM 6 (Pickert **(9)**). *If the plane is* $B - (x = 0)$ *transitive, then* (i) *multiplication is associative and* (ii) *the coordinatization is linear.*

In each of the remaining theorems of this section, it is to be understood that the plane is $p_1 - L_1$ transitive and also $p_2 - L_2$ transitive. It is also to be understood that p_1 is different from p_2 and L_1 is different from L_2 unless equality is specifically indicated.

THEOREM 7. *If* $L_1 = L_2 = L$ *and both* p_1 *and* p_2 *are on L, the plane is a Veblen-Wedderburn plane with L as the line at infinity.*

Proof. By **(4**, Lemma 3.2**)**, the plane is $p - L$ transitive for every point p on L. This implies that the plane is a Veblen-Wedderburn plane. See **(1)**.

THEOREM 8. *If* p_1 *is on* L_1 *and* p_2 *is the intersection of* L_1 *and* L_2, *the plane can be coordinatized by a distributive quasifield.*

Proof. By the use of the $p_2 - L_2$ transitivity, p_1 can be carried into any point on L_1 except p_2. Hence, by Theorem 7, we have a Veblen-Wedderburn plane with L_1 as L_∞. Theorem 8 then follows from Theorem 4.

THEOREM 9. *If $p_1 = p_2 = p$, L_1 goes through p and L_2 does not go through p, then the plane can be coordinatized by a nearfield.*

Proof. Take $p \equiv A$, $L_1 \equiv L_\infty$ and $L_2 \equiv (y = 0)$. Then, by Theorem 5, the coordinatization is linear and addition is a group. Now consider the $A - (y=0)$ perspectivity which carries the point $(0, 1)$ into $(0, a)$. Since B is fixed, $y = 1 \rightarrow y = a$. Since every line through A is fixed, the x coordinate of every point is unchanged. In particular,

$$(1, 1) \rightarrow (1, a).$$

Hence $y = x \rightarrow y = xa$. It follows that $(c, c) \rightarrow (c, ca)$ and $y = c \rightarrow y = ca$. Thus, the right multiplications are isomorphic to the group of $A - (y = 0)$ perspectivities, and multiplication by non-zero elements is associative. Now, if (1) indicates the point at infinity corresponding to slope 1 and (a) the point at infinity corresponding to slope a, we have

$$(1) \rightarrow (a); \text{ moreover, } (0, b) \rightarrow (0, ba).$$

Hence, $y = x + b \rightarrow y = xa + ba$. But $(c, c + b) \rightarrow (c, (c + b)\, a)$. The latter point must be on the line $y = xa + ba$. Hence $(c + b)a = ca + ba$.

André **(1**, Theorem 13**)** has shown that addition is abelian if all of the other properties of a Veblen-Wedderburn system are satisfied. Properties 3, 4, 5 and 6 of a Veblen-Wedderburn system follow from the general properties of Hall's ternary and the theorem follows. Baer **(4)** has given the proof for the dual of this case.

THEOREM 10. *If p_1 is on both L_1 and L_2 and p_2 is on L_2 and if we take $p_1 \equiv A_1$, $p_2 \equiv B$, $L_2 \equiv (x = 0)$, we have* (i) *the coordinatization is linear* (ii) *addition forms a group and* (iii) *multiplication is associative.*

Proof. Follows from Theorems 5 and 6.

THEOREM 11. *If p_1 and p_2 are on L_1, but L_2 does not go through p_1 or p_2, then the plane can be coordinatized by a nearfield.*

Proof. Since p_1 is not fixed by the $p_2 - L_2$ perspectivities, there are other points p on L_1 which are centers of $p - L_1$ transitivities. It follows from Theorem 7 that the plane is a Veblen-Wedderburn plane with L_1 as L_∞. If we take $p_2 \equiv B$ and $L_2 \equiv (x = 0)$, multiplication is associative by Theorem 6.

THEOREM 12. *If p_1 is on L_1, neither p_1 nor p_2 is on L_2, and p_2 is not on L_2, then the plane is Desarguesian.*

Proof. By the $p_1 - L_1$ perspectivities, p_2 can be carried into any point on the line $p_1 p_2$ except p_1. By the $p_2 - L_2$ perspectivities, p_1 can be carried into any point on $p_1 p_2$, excepting p_2 and the point on L_2. It follows that the group of collineations is transitive on points of $p_1 p_2$ and every point on $p_1 p_2$ is the center of two $p - L$ transitive groups. One of these is a group of

elations and the other is a group of homologies. If we take any point on $p_1 p_2$ as the point p in Theorem 9, we see that we have a nearfield plane. If r is the intersection of L_1 and L_2, the line pr will be our line at infinity. The existence of collineations moving the line at infinity implies the other distributive law **(1**, Theorem 18**)**.

THEOREM 13. *If $L_1 = L_2 = L$ and neither p_1 nor p_2 is on L, the plane can be coordinatized by a dual nearfield.*

Proof. Any point on the line $p_1 p_2$ (excepting the point on L) can be carried into any other point. André **(2**, Theorem 3**)** has shown that the group of elations with axis L is transitive on these points. That is, if q is the intersection of $p_1 p_2$ with L, the plane is $q - L$ transitive. Our theorem then follows from the dual of Theorem 9.

THEOREM 14. *If p_1 is on L_2 and p_2 is on L_1, but p_1 is not on L_1 nor p_2 on L_2, then the plane can be coordinatized by a system in which*

(i) *the coordinatization is linear,*
(ii) *multiplication is associative,*
(iii) $a(c + b) = ac + ab$.

Proof. Take L_1 as L_∞, p_1 as $(0, 0)$, p_2 as B and L_2 as $x = 0$. Then (i) and (ii) are established by Theorem 6. Now, consider the $p_1 - L_1$ perspectivity which carries the line $x = 1$ into $x = a$. Since $(0, 0)$ is the center, each line $y = xm$ is fixed, so that $(1, m) \to (a, am)$. Thus, the line $y = m$ maps into $y = am$. But, since the line $y = x$ is fixed, $(m, m) \to (am, am)$ and the line $x = m \to x = am$. Now

$$y = x + b \to y = x + ab.$$

Therefore $(c, c + b) \to (ac, ac + ab)$. But $y = c + b \to y = a(c + b)$. Hence $a(c + b) = ac + ab$.

THEOREM 15. *If p_1, p_2, p_3 are the vertices of a triangle in which L_1, L_2, L_3 are the respective opposite sides and if the plane is $p - L$ transitive with respect to each vertex and its opposite side, the plane can be coordinatized by a neofield,* i.e.,

(i) $x \cdot m \circ b = xm + b$,
(ii) *multiplication is associative,*
(iii) $a(c + b) = ac + ab$,
(iv) $(b + c)\, a = ba + ca$.

Proof. Take p_1 as $(0, 0)$, p_2 as B, p_3 as A. By Theorem 14, we have only to prove (iv). In the group of perspectivities with center A and axis $y = 0$, consider the perspectivity which carries $y = 1$ into $y = a$. Every line $x = c$ is fixed, so $(c, 1) \to (c, a)$. Now, $(0, 0)$ is fixed, so $y = xc \to y = xm$ for some m. The point $(c, 1)$ is on $y = xc^{-1}$, so (c, a) is on $y = xm$. Thus $a = cm$ and $m = c^{-1}a$ so that $(c^{-1}) \to (c^{-1}a)$.

In particular, consider $y = x + c$, which goes through the point (1) and the fixed point $(-c, 0)$. Its image must go through (a) and $(-c, 0)$. Thus

$$y = x + c \to y = xa + ca.$$

Therefore $(b, b + c) \to (b, ba + ca)$. But $y = b + c \to y = (b + c)a$. Hence $ba + ca = (b + c)a$.

We note here that Hughes **(6)** has used different groups of collineations to obtain coordinatinizations like those in Theorems 10, 14, and 15.

THEOREM 16. *If p_1 is not on L_1 and p_2 is not on L_2 and either* (i) *p_2 is on L_1 but p_1 is not on L_2 or* (ii) *p_1 is not on L_2, p_2 is not on L_1 and p_1 and p_2 are not collinear with the intersection r of L_1 with L_2, then the plane is Desarguesian.*

Proof. (i) Note that L_1 is fixed by both the $p_1 - L_1$ perspectivities and the $p_2 - L_2$ perspectivities. The points on $p_1 p_2$ (excepting p_2) are transitive under collineations which leave L_1 fixed. Hence each point $p \neq p_2$ on $p_1 p_2$ is the center of a $p - L_1$ transitive group. This implies **(2**, Theorem 3**)** that the plane is $p_1 - L_2$ transitive.

By Theorem 9, if we take $p_2 \equiv A$, $L_1 \equiv L_\infty$, $L_2 \equiv (y = 0)$ the plane will be coordinatized by a nearfield. If we take the line $p_1 p_2$ as $x = 0$, the plane will be $(0, 0) - L_\infty$ transitive. By methods similar to those which have been used in other proofs, it can be shown that the $(0, 0) - L_\infty$ perspectivity which carries $(1, 1)$ into (a, a) carries $y = c$ into $y = ac$ and $x = c$ into $x = ac$ and finally, that $a(c + b) = ac + ab$. Hence, the coordinatizing algebra is a field.

(ii) If neither p_1 nor p_2 is on L_1 or L_2 and r is the intersection of L_1 and L_2, the $p_1 - L_1$ and $p_2 - L_2$ perspectivities generate a group which is doubly transitive on the points of $p_1 p_2$ and leaves r fixed. Hence, for every pair of points q and s on $p_1 p_2$ $(q \neq s)$, the plane is $q - sr$ transitive. Hence, keeping s fixed and varying q, the plane is $s - sr$ transitive for each s on $p_1 p_2$ **(2**, Theorem 3**)**. Taking s as the point p in Theorem 9, the plane can be coordinatized by a nearfield with sr as L_∞. But sr is not fixed, so the nearfield must be a field.

The above theorems characterize the various types of planes uniquely in the sense that any plane coordinatized by the algebra appropriate to each theorem will have the corresponding $p - L$ transitivities. If we include the cases dual to those considered here, we have taken care of all cases of pairs of $p - L$ transitivities except the following two:

I. Neither p_1 nor p_2 is on L_1 or L_2 but the line $p_1 p_2$ goes through the intersection of L_1 and L_2.

II. p_1 is on L_1 and p_2 is on L_2 with $p_1 \neq p_2$ and $L_1 \neq L_2$.

3. Multiple transitivity.

THEOREM 17. *Let π be a finite projective plane with $n + 1$ points on a line. Then if n is an odd power of two and the plane is doubly transitive, π is Desarguesian.*

Proof. If π is doubly transitive and not of square order, there is a perspectivity of period two (involution) with some center p and axis L, where $p \in L$ for n even **(7)**. If $q \notin L$ and $q_1 \neq L$ the existence of a collineation which fixes p and carries q into q_1 implies the existence of an involution with center p, axis pq_1. Thus every line through p is an axis of a perspectivity of period two with center p. Now, by Lemma 2, the product of two elations with the same center p is an elation with center p. By Theorem 3, all elations with the same center are of the same period. Hence the product of two involutions with center p is an involution with center p, all axes going through p, since n is even.

Thus the involutions with center p form a group whose order is a power of 2. The involutions with center p and any given axis through p form a subgroup. Let 2^r be the order of the group with center p, 2^s the order of the group with center p, any given axis. Then, since there are $n + 1$ lines through p,

$$(n + 1)(2^s - 1) + 1 = 2^r, 2^s n + 2^s - n = 2^r,$$

where $r > s$. Thus 2^s must divide n. Suppose $n = 2^{s+t}$. Then

$$2^{2s+t} + 2^s - 2^{t+s} = 2^r, 2^{s+t} + 1 - 2^t = 2^{r-s}.$$

Now, since $r > s$, the right-hand side is even and the left-hand side is odd unless $t = 0$. If such be the case, $n = 2^s$. Thus the order of the group of involutions with center p, fixed axis through p, is equal to n. Hence we have $p - L$ transitivity for every line L through p. This is true for every point in the plane. This implies projective satisfaction of the Minor Theorem of Desargues. Since the plane is finite, it must then be Desarguesian.

THEOREM 18. *If n is an odd power of two, and the plane is doubly transitive for points not on the line L_∞, it is a Veblen-Wedderburn plane.*

Proof. Unless L_∞ is fixed by all collineations, the plane is doubly transitive for all points and we can apply Theorem 17. Suppose L_∞ is fixed by every collineation of the plane. Now the double transitivity implies the existence of perspectivities of period 2. If L_∞ is the axis of one of these involutions, any two points can be interchanged by an involution with L_∞ as axis. Thus finite points are transitive under elations with L_∞ as axis ("translations"), and the plane is Veblen-Wedderburn **(1)**.

If L_∞ is not the axis of any of these involutions, their centers must be on L_∞ if L_∞ is to be fixed by all collineations. The plane will be transitive on finite lines, so every finite line is the axis of an involution. Let p be any point on L and let q be any finite point. Each involution which interchanges q with a point on the line pq must have p as center. (The double transitivity implies that every pair of finite points can be interchanged by a perspectivity of order two.) Moreover, p is the center of the involution with axis pq, since our involutions are elations. Corresponding to the n finite lines through p, there are at least n involutions with center p.

Now, let $r \neq p$ be some point on L_∞. Since r has only n possible images (including itself) under involutions with center p, either r is fixed by some involution with center p or at least two of them interchange r with the same point. We are dealing with elations, so that the only fixed points are on the axes. Thus if r is fixed by some involution, L_∞ is the axis of that involution. We have already noted that the product of two involutions with center p is an involution with center p. In the case where two involutions interchange r with the same point, the product of these two involutions is an involution fixing r. In every case, we are led to an involution with L_∞ as axis. We have already remarked that this is sufficient to establish the theorem.

Now, every collineation of period two either is a perspectivity or leaves fixed all of the points of a subplane of order $n^{\frac{1}{2}}$. The author's proof that doubly transitive planes are Desarguesian fails only in the case that no collineation of period two is perspectivity. Thus (even if n is a square) if n is odd or a power of two and the plane is doubly transitive, it is Desarguesian unless every collineation of period two leaves a subplane of order $n^{\frac{1}{2}}$ pointwise fixed.

Definition. A plane which contains no proper subplanes will be called a *minimal plane.*

THEOREM 19. *If the plane π is finite and is transitive on proper quadrangles, then every quadrangle generates a Desarguesian subplane of prime order μ, where μ divides n or $n - 1$.*

Proof. If π is finite, it certainly contains minimal subplanes. The key to our argument is the fact that two minimal subplanes cannot share a quadrangle. Let π_2 be a minimal subplane of π. Any collineation of π which carries a quadrangle of π_1 into a quadrangle of π_2 is also a collineation of π_1. Hence π_1 is transitive on quadrangles and is doubly transitive. If the order of π_1 is odd, π_1 is Desarguesian. If the order μ of π_1 is even, π_1 will be Desarguesian if μ is a power of 2. Now π_1 admits perspectivities of period two. But if μ is even this implies that π_2 contains a Fano subplane **(8)**. Thus if μ is even, μ is equal to 2.

Clearly, every quadrangle will generate a subplane of order μ and since π_1 is a minimal Desarguesian plane, μ is a prime.

Now consider three fixed points p, q, and r on a line L and a line $L_1 \neq L$ through p. Let $s \neq p$ be a fixed point on L_1. There will be $n - 1$ quadrangles such that q and r are diagonal points, s is one vertex and L_1 is one side. The points p, q, and r will be in the minimal subplane generated by any one of these quadrangles. These quadrangles will occur in subsets with $\mu - 1$ in each subset, those in the same subset belonging to the same minimal subplane. Hence $\mu - 1$ divides $n - 1$. Similarly, if we allow s to vary, $\frac{1}{2}\mu(\mu - 1)$ divides $\frac{1}{2}n(n - 1)$. Since μ is prime, μ divides n or $n - 1$.

Furthermore, there are $(n^2 + n + 1)\ (n^2 + n)\ n^2\ (n - 1)^2$ ordered quadrangles in π and

$$(\mu^2 + \mu + 1)\ (\mu^2 + \mu)\ \mu^2\ (\mu - 1)^2$$

ordered quadrangles in each minimal subplane. Hence

$$(\mu^2 + \mu + 1)\,(\mu^2 + \mu)\,\mu^2\,(\mu - 1)^2$$

divides $(n^2 + n + 1)\,(n^2 + n)\,n^2\,(n - 1)^2$. If we consider the class of quadrangles containing some fixed point, we get by a similar argument that $(\mu^2 + \mu)\,\mu^2\,(\mu - 1)^2$ divides $(n^2 + n)\,n^2\,(n - 1)^2$.

4. The existence of Desarguesian configurations.

THEOREM 20. *In a finite projective plane, let L_1, L_2 and L_3 be three distinct lines through a point p. Let r and s be any two points not on L_1, L_2 or L_3. Consider the set of triangles with one vertex each on L_1, L_2, and L_3, one side going through r and the other side going through s. At least one pair of triangles of this set satisfies Desargues's Theorem.*

Proof. Let L_1, L_2, and L_3 respectively intersect the line rs in the points t_1, t_2, and t_3. Let A_1, A_2, and A_3 be the vertices of a triangle such that A_1 is on L_1, A_2 is on L_2, and A_3 is on L_3. Let these points be chosen so that the line A_2A_3 goes through r and A_2A_1 goes through s. A_2 may be any point on L_2 except p or t_2. Hence there are $n - 1$ such triangles.

Now the line A_1A_3 cannot intersect rs in t_3 or t_2. Neither can A_1A_3 go through r or s. Hence there are at most $(n + 1) - 4 = n - 3$ possible points in which A_1A_3 can intersect rs. Since there are $n - 1$ such triangles, there must be a pair of triangles $A_1A_2A_3$ and $A_1'A_2'A_3'$ such that A_1A_3 and $A_1'A_3'$ intersect in a point on rs. The triangles $A_1A_2A_3$ and $A_1'A_2'A_3'$ then satisfy Desargues's Theorem.

A similar theorem can be proved for $r = t_2$ $s = t_1$ if Fano's configuration does not exist in the plane.

Added in proof. Since submitting this paper, the author has received communications from A. Wagner which make Theorems 17 and 19 outmoded. Wagner has shown that (1) double transitivity implies that the plane is Desarguesian provided only that n is even, and (2) without regard to restrictions on n, transitivity on quadrangles implies that the plane is Desarguesian.

References

1. J. André, *Ueber nicht-Desarguesche Ebene mit transitiver Translationsgruppe*, Math. Z., *60* (1954), 156–186.
2. ———, *Ueber Perspektivitäten in endlichen projektiven Ebenen*, Arch. Math., *6* (1955), 29–32.
3. ———, *Projektive Ebenen über Fastkörpern*, Math. Z., *62* (1955), 137–160.
4. R. Baer, *Homogeneity of projective planes*, Amer. J. Math., *64* (1942), 137–152.
5. M. Hall, *Projective planes*, Trans. Amer. Math. Soc., *54* (1943), 229–277.
6. D. R. Hughes, *Partial difference sets*, Amer. J. Math., *78* (1956), 650–674.
7. T. G. Ostrom, *Double transitivity in finite projective planes*, Can. J. Math., *8* (1956), 563–567.
8. ———, *Ovals, dualities, and Desargues's theorem*, Can. J. Math., *7* (1955), 417–431.
9. G. Pickert, *Projektive Ebenen* (Berlin, 1955).

Montana State University

DEDEKIND COMPLETENESS AND A FIXED-POINT THEOREM

E. S. WOLK

1. Introduction. McShane **(5, 6)** has introduced the concept of "Dedekind completeness" for partially ordered sets, which seems to be a natural generalization of the usual concept of completeness for lattices. It is the purpose of this paper to discuss some of the properties of Dedekind completeness, particularly with respect to a rather natural class of partially ordered sets which we call "uniform." Among our results we obtain an analogue of MacNeille's "completion by cuts." We also extend the well-known fixed-point theorem, due to Tarski **(7)**, and then generalize the characterization of a complete lattice due to Davis **(3)**.

2. Dedekind completeness. Let P be a partially ordered set (poset) with respect to a relation $\leqslant$. We assume that P has a greatest element I and a least element O.

DEFINITION 1. We say that a set $S \subset P$ is *up-directed* if and only if for each $a \in S$, $b \in S$, there exists $c \in S$ with $a \leqslant c$, $b \leqslant c$. Dually, S is *down-directed* if and only if for each $a \in S$, $b \in S$, there exists $c \in S$ with $c \leqslant a$, $c \leqslant b$.

Thus, any subset of P which has a greatest element is up-directed, and dually. The following definition is essentially that of McShane.

DEFINITION 2. A poset P is *Dedekind complete* if and only if every up-directed subset of P has a least upper bound in P and every down-directed subset has a greatest lower bound in P.

Example 1. It is clear that the concepts of Dedekind completeness and ordinary completeness coincide if P is a lattice. A simple example of a Dedekind complete poset, which is not a lattice, is provided by the set C of all closed disks in the Euclidean plane E_2, partially ordered by set inclusion, and with O and I elements adjoined. To show that C is Dedekind complete, let A be an up-directed subset of C, and let

$$X = \{x | x \in E_2 \text{ and } x \in a \text{ for some } a \in A\}.$$

If X is an unbounded subset of E_2, then clearly l. u. b. $(A) = I$. If X is bounded choose two points x and y in the closure of X such that the distance from x to y is equal to the diameter of X. Let m be a closed disk with the line seg-

Received January 12, 1956.

ment connecting x and y as its diameter. Straightforward arguments then show that

(i) no point of X is exterior to m, and

(ii) every interior point of m is a point of X. Thus m = l. u. b. (A). The obvious dual argument will then show that any down-directed subset of C has a g. l. b.

If $A \subseteq P$, let

$$A^* = \{x | x \in P \text{ and } x \geqslant a \text{ for all } a \in A\},$$
$$A^+ = \{x | x \in P \text{ and } x \leqslant a \text{ for all } a \in A\}.$$

We shall write A^{*+} for the set $(A^*)^+$. We shall make important use of the following concept:

DEFINITION 3. A poset P is *uniform* if and only if A^* is a down-directed set for every up-directed subset A, and dually, B^+ is up-directed for every down-directed subset B.

Any lattice is obviously a uniform poset. As an example of a uniform poset, which is not a lattice and not Dedekind complete, we may take the set of all closed disks in the plane with rational radii, partially ordered by set inclusion, and with O and I elements adjoined.

We have the following trivial lemma:

LEMMA 1. *A uniform poset P is Dedekind complete if and only if every up-directed subset of P has a* l.u.b. *in P (or every down-directed subset of P has a* g. l. b. *in P).*

We shall also use a strong form of Zorn's lemma due to Bourbaki **(2)**:

LEMMA (Bourbaki). *If every well-ordered chain in a poset S has an upper bound in S, then S has a maximal element.*

As a consequence of the above lemma the reader may easily deduce

LEMMA 2. *If Z is any chain in a poset P, then there exists a well-ordered chain $C \subset Z$ with $C^* = Z^*$.*

We now have the following theorem:

THEOREM 1. *A poset P is Dedekind complete if and only if P is uniform and every well-ordered chain in P has a* l. u. b.

Proof. If P is Dedekind complete, and S is up-directed in P, then S^* has a least element and hence is down-directed. The obvious dual statement also holds: thus P is uniform and the conclusion follows. Conversely, let P be uniform and suppose that every well-ordered chain in P has a l. u. b. Let A be any down-directed subset of P, and let Z be a maximal chain in A^+. We assert that Z has a l. u. b., m, for otherwise Lemma 2 would provide us with a contradiction of our hypothesis. If $a \in A$, we have $a \geqslant z$ for all $z \in Z$;

hence $a \geqslant m$ and $m \in A^+$. By maximality of Z, m is a maximal element of A^+. We assert that m is the greatest element of A^+. For suppose that there exists $c \in A^+$ with $c > m$. Since A^+ is up-directed, there exists $x \in A^+$ with $x \geqslant m$, $x \geqslant c$, contradicting the maximality of m. Thus $m = \text{g. l. b. } (A)$, and P is Dedekind complete by Lemma 1.

As a corollary we have the following known result, for which a proof seems to have thus far been lacking in the literature:

COROLLARY. *A lattice L is complete if and only if every well-ordered chain in L has a* l. u. b.

Let us call a chain Z in P *inversely well-ordered* if and only if every subset of Z has a greatest element. We then have obvious dual formulations of Lemma 2 and Theorem 1. We shall also need the following lemma, which extends a result of Davis **(3**, Lemma 1, p. 311**)**; our proof of it becomes trivial by employing Zorn's lemma (rather than transfinite induction as in **(3)**):

LEMMA 3. *Let P be a uniform poset, and let Z be an inversely well-ordered chain in P with no* g. l. b. *in P. Then there exists a well-ordered chain Y in P such that*

(i) *$y \in Y$ implies $y < z$ for all $z \in Z$, and*
(ii) *$Y^* \cap Z^+$ is empty.*

Proof. Z^+ is up-directed, by our hypothesis of uniformity; hence Z^+ has no maximal elements. Then by the lemma of Bourbaki there exists a well-ordered chain Y in Z^+ such that $Y^* \cap Z^+$ is empty.

3. Imbedding of a uniform poset in a Dedekind complete poset. We shall now obtain an analogue of MacNeille's well-known imbedding of a poset in a complete lattice **(4**; also see **1**, p. 58**)**.

DEFINITION 4. A subset J of a poset P is a *normal ideal* in P ("closed ideal" in the terminology of Birkhoff) if and only if $J^{*+} = J$. A subset of P of the form

$$J_a = \{x | x \in P \text{ and } x \leqslant a\}$$

is called a *principal ideal*.

LEMMA 4. *A subset of P is a normal ideal if and only if it is the intersection of a set of principal ideals* (cf. **1**; p. 62, problem 4**)**.

Proof. Let $S \subset P$ and let

$$A = \bigcap_{x \in S} J_x.$$

Then $A = S^+$. In general we have $S \subset S^{+*}$; hence $S^+ \supset (S^{+*})^+$, or $A \supset A^{*+}$. Since in general $A \subset A^{*+}$, it follows that A is a normal ideal. Conversely, if A is a normal ideal in P, then

$$A = (A^*)^+ = \bigcap_{x \in A^*} J_x$$

For uniform posets we now have another characterization of Dedekind completeness, which generalizes a known result for complete lattices (**1**; p. 59, exercise 2):

THEOREM 2. *A uniform poset P is Dedekind complete if and only if every up-directed normal ideal in P is principal.*

Proof. Let P be Dedekind complete, and let J be an up-directed normal ideal in P. Then J has a l. u. b. m, and $m \in J^{*+} = J$. It follows that J is principal. To prove the converse, let A be a down-directed subset of P. By Lemma 4,

$$A^+ = \bigcap_{a \in A} J_a$$

is a normal ideal, which by hypothesis is up-directed. Hence A^+ has a l. u. b., which is the g. l. b. of A. Thus P is Dedekind complete by Lemma 1.

Now let $N(P)$ be the set of all up-directed normal ideals of P, partially ordered by inclusion. The correspondence $x \leftrightarrow J_x$ is a one-to-one order-preserving mapping of P into a subset of $N(P)$. Furthermore, we have

THEOREM 3. *If P is a uniform poset, then $N(P)$ is Dedekind complete.*

Proof. Let Σ be an up-directed subset of $N(P)$, and let

$$A = \bigcup_{J \in \Sigma} J$$

(where $\bigcup$ denotes set union). It is easily seen that A is an up-directed subset of P. Hence A^* is down-directed, and A^{*+} is up-directed. Since A^{*+} is the smallest normal ideal containing A, we have A^{*+} = l. u. b. (Σ). Now let Ω be a down-directed subset of $N(P)$. We first show that

$$B = \bigcup_{J \in \Omega} J^*$$

is a down-directed subset of P. Let a and b be arbitrary elements of B; then there exist $J_1, J_2 \in \Omega$ with $a \in J_1^*$, $b \in J_2^*$. By our hypothesis on Ω, there exists $J_3 \in \Omega$ with $J_3 \subset J_1 \cap J_2$. Then $J_3^* \supset (J_1 \cap J_2)^* \supset J_1^* \cup J_2^*$. But J_3^* is down-directed, by uniformity of P: hence there exists $c \in J_3^*$ with $c \leqslant a$, $c \leqslant b$, and thus B is down-directed. Now let

$$K = \bigcap_{J \in \Omega} J.$$

But

$$\bigcap_{J \in \Omega} J = \bigcap_{J \in \Omega} J^{*+} = \left(\bigcup_{J \in \Omega} J^*\right)^+ = B^+.$$

Hence K is an up-directed normal ideal, and K = g. l. b. (Ω).

Example 2. Let P be the set of all closed disks in the plane with rational radii, ordered by inclusion. If z is an arbitrary closed disk in the plane, then the set

$$S(z) = \{a | a \in P \text{ and } a \subset z\}$$

is an up-directed normal ideal in P; and conversely, the reader may verify that every such ideal is of the form $S(z)$ for some disk z. Hence the "Dedekind completion" $N(P)$ is isomorphic to the set of all closed disks in the plane.

Example 3. If P is not uniform, then $N(P)$ may fail to be Dedekind complete. We construct an example of such a poset P as follows. Let $A = \{a_{ij}\}$ $(i = 1, 2, \ldots; j = 1, 2, \ldots)$ be an infinite rectangular array, in which i denotes the column index, j the row index. We partially order A by defining $a_{ij} < a_{mn}$ if and only if $i < m$ or $j < n$. We surmount this array with a sequence $\{z_i\}$ of mutually incomparable elements (with respect to our ordering) such that $a_{ij} < z_i$ for each i and each j. We adjoin two more incomparable elements x and y which are upper bounds for the set $\{z_i\}$. We then let P be the set consisting of the array $A = \{a_{ij}\}$, the set $\{z_i\}$, the elements x and y, and O and I elements; and let P be partially ordered as described above. Thus we have $a_{ij} < z_k$ if and only if $i \leqslant k$. We see that A is an up-directed subset of P. (Note, however, that A^{*+} is the union of A and the set $\{z_i\}$, and hence is not up-directed. Thus A^{*+} is not an element of $N(P)$). Hence

$$\Sigma = \{J_a | a \in A\}$$

is an up-directed subset of $N(P)$. But the set Σ^* contains J_x and J_y as minimal elements; hence Σ has no l. u. b. in $N(P)$.

4. The fixed-point theorem. If f is a function mapping a poset P into itself, we say that f is *isotone* if and only if $x \leqslant y$ implies $f(x) \leqslant f(y)$. x is a *fixed-point* of f if and only if $x = f(x)$. For any isotone function f on P let us write $H(f) = \{x | x \in P \text{ and } x \leqslant f(x)\}$.

DEFINITION 5. An isotone function f on a poset P is *directable* if and only if $H(f)$ is an up-directed subset of P.

The reader may verify that any isotone function on a lattice is directable. Thus the following theorem generalizes the fixed-point theorem of Tarski (7, Theorem 1):

THEOREM 4. *If every up-directed subset of a poset P has a* l. u. b. *in P, then every directable function on P has a fixed-point.*

Proof. Let f be a directable function on P and let $u =$ l. u. b. $[H(f)]$. We easily prove, precisely as in the proof of Theorem 1 of (7), that u is a fixed-point of f. We omit the details.

We now obtain a generalization of the result of Davis (3, Theorem 2):

THEOREM 5. *If every directable function on a uniform poset P has a fixed-point, then P is Dedekind complete.*

Proof. Assume P is not Dedekind complete. Applying the dual formulation of Theorem 1 and then Lemma 3, we infer that there exist two chains Y and Z in P such that

(i) Y is well-ordered and Z is inversely well-ordered,
(ii) $y \in Y$ implies $y < z$ for all $z \in Z$,
(iii) $Y^* \cap Z^+$ is empty.

We shall proceed to obtain a contradiction by defining a directable function f on P which has no fixed-points. We do this exactly as in **(3, pp. 313-314)**. To define $f(x_0)$ for an arbitrary $x_0 \in P$ we distinguish two cases:

$$(1)\ x_0 \in Z^+, \qquad (2)\ x_0 \notin Z^+.$$

In case (1) we have $x_0 \notin Y^*$. Let $Y(x_0) = \{y | y \in Y \text{ and } y > x_0\}$. $Y(x_0)$ has a least element y_0, which we define as $f(x_0)$. In case (2), let $Z(x_0) = \{z | z \in Z \text{ and } z < x_0\}$. $Z(x_0)$ has a greatest element z_0, which we define as $f(x_0)$. It is clear that f can have no fixed-points. The proof that f is isotone is identical with that in **(3, p. 314)**: we therefore omit the details. It remains to show that f is directable. From our definition of f it is clear that $x \in H(f)$ implies that f falls in case (1) above; i.e., $x \in Z^+$. Also it is clear that $Y \subset H(f)$. Now suppose that we have $a \in H(f)$, $b \in H(f)$. Then $a < f(a)$, $b < f(b)$, and $f(a) \in Y$, $f(b) \in Y$. Let $c = \max \{f(a), f(b)\}$. We have $c > a$, $c > b$, and $c \in H(f)$, thus completing the proof.

Combining Theorems 4 and 5, we obtain the following characterization of Dedekind completeness:

COROLLARY. *A uniform poset P is Dedekind complete if and only if every directable function on P has a fixed-point.*

REFERENCES

1. G. Birkhoff, *Lattice theory*, Amer. Math. Soc. Colloquium Publications, *25* (1948).
2. N. Bourbaki, *Sur le théorème de Zorn*, Archiv. der Math., *6* (1949–50), 434–437.
3. Anne C. Davis, *A characterization of complete lattices*, Pacific J. Math., *5* (1955), 311–319.
4. H. MacNeille, *Partially ordered sets*, Trans. Amer. Math. Soc., *42* (1937), 416–460.
5. E. J. McShane, *Order-preserving maps and integration processes*, Annals of Math. Studies, *31* (Princeton, 1953).
6. E. J. McShane, *Partial orderings and Moore-Smith limits*, Amer. Math. Monthly, *59* (1952), 1–11.
7. A. Tarski, *A lattice-theoretical fixpoint theorem and its applications*, Pacific J. Math. *5* (1955), 285–309.

University of Connecticut

CYCLICALLY ORDERED SETS

M. I. AISSEN

Introduction. The classical separation theorems regarding the relative position of the zeros of two consecutive orthogonal polynomials, or of the positive zeros of two consecutive Bessel Functions, are well known. Less well known is a theorem of Bôcher and Porter **(1**, p. 207**)**, which states that the positive zeros of three consecutive Bessel Functions occur in cyclic order. This paper extends the result to symmetric orthogonal polynomials and more generally to sequences of functions satisfying a certain type of difference equation. The investigation arose in an attempt to prove a conjecture of Schoenberg **(3)** for which it was believed that knowledge of the zeros of polynomials generated by $\exp(x + tx^k)$ would be helpful. This conjecture was proved by Edrei **(2)**.

1. Notation. By $Z(A)$ we mean the number of elements in the set A. If A is not finite, we do not distinguish among different cardinals, and we write $Z(A) = \infty$. The elements of the various sets discussed are real numbers. Any term involving order such as "consecutive," "least," etc. refers to the natural order of the real numbers. We denote the empty set by ϕ.

If I is an interval, then we denote $Z(A \cap I)$ by $Z_I(A)$. The symbols I, J, I', I_n, J^* always refer to intervals. The term interval is used in the wide sense and includes those which are open or closed at either end and which may be finite, semi-infinite, or infinite. The symbol k always refers to some fixed but unspecified integer greater than one.

2.Cyclically ordered sets. Let $a_1 < a_2 < \ldots < a_N$ be N real numbers. We consider the sets E_μ $\mu = 1, 2 \ldots k$, such that $a_m \in E_\mu$ if and only if $m \equiv \mu \pmod{k}$. The finite sets E_μ are examples of cyclically ordered sets. If $0 \leqslant N < k$, some or all of the sets will be empty. We state without proof some of the properties of the sets E_μ. Let E denote $\bigcup_\mu E_\mu$:

(i) $E_\mu \cap E_\lambda = \phi$, $\lambda \neq \mu$

(ii) Any k consecutive elements of E include one from each of the sets E_μ.

(iii) If $Z(E_\lambda) = 0$ for some λ, $Z(E_\mu) \leqslant 1$ for all μ.

(iv) $|Z(E_\lambda) - Z(E_\mu)| \leqslant 1$ for all λ and μ.

(v) Between any two elements of E_λ there is an element of E_μ if $\lambda \neq \mu$.

(vi) If $E^*_\mu = E_\mu \cap I$, where I is a fixed interval and $E^* = E \cap I = \bigcup_\mu E_\mu^*$, then the sets E^*_μ and E^* satisfy (i)–(v).

Received April 16, 1956.

These properties are not independent. For example, (iii) is a consequence of either (iv) or (v). On the other hand (i), (ii), and (iii) imply (iv), (v), and (vi). We now formulate the following definition:

Definition. We say that the k sets $E_1, E_2, \ldots E_k$ are *cyclically ordered* (*C-O*) if and only if they satisfy the following conditions

C1: $E_\lambda \cap E_\mu = \phi$, $\lambda \neq \mu$.

C2: $Z_J(E_\lambda) \leqslant Z_J(E_\mu) + 1$ for all λ, μ and an arbitrary interval J.

THEOREM 1. *Let $E_1, E_2, \ldots, E_k$ be C-O. Let $E^*_\mu = E_\mu \cap J$ where J is a fixed interval. Then $E^*_1, E^*_2, \ldots, E^*_k$ are C-O.*

Proof. C1 is satisfied since $E^*_\lambda \cap E^*_\mu = E_\lambda \cap E_\mu \cap J$.

C2 is satisfied since $Z_I(E^*_\lambda) = Z_I{}^*(E_\lambda)$ where for an arbitrary interval, I, $I^* = I \cap J$.

THEOREM 2. *A necessary and sufficient condition that the sets $E_1, E_2, \ldots, E_k$ be C-O, is that each pair of sets E_λ, E_μ $(\lambda \neq \mu)$ be C-O.*

Proof. The proof is immediate since C1 and C2 are required to hold for all pairs of sets E_λ, E_μ.

THEOREM 3. *Let $E_1, E_2, \ldots, E_k$ be C-O and let $E = \bigcup_\mu E_\mu$. Then*

$$Z_I(E) \leqslant k \Rightarrow Z_I(E_\mu) \leqslant 1,$$

for all μ. In particular $Z_I(E) = k, \Rightarrow Z_I(E_\mu) = 1$ for all μ.

Proof. By C1,

$$Z_I(E) = \sum_\mu Z_I(E_\mu). \quad Z_I(E_\lambda) > 1 \Rightarrow Z_I(E_\mu) \geqslant 1$$

for all μ by C2. Hence $Z_I(E_\lambda) > 1 \Rightarrow Z_I(E) > k$ and thus $Z_I(E) \leqslant k \Rightarrow Z_I(E_\mu) \leqslant 1$, all μ.

Further, if $Z_I(E_\lambda) = 0$ for some λ, $Z_I(E) < k$. Hence $Z_I(E) = k \Rightarrow Z_I(E_\mu) = 1$ for all μ.

3. The principal theorem. Historically, cyclically ordered sets have been studied only in the case, $k = 2$ (except for the theorem of Bôcher-Porter), in which case the sets involved are said to be separated. In many cases the sets are zeros of appropriate continuous functions which are related by a linear difference equation of the second order. The functional equation is used to determine the sign of the functions at critical points and the Darboux property of continuous functions is then invoked to show that one function vanishes in some interval. It is rather curious that in such cases the method is used only to show that pairs of sets are *C-O*, since it can usually be used to show that three sets are *C-O*. This can be generalized to the case of certain kth order linear difference equations in which case $k + 1$ sets will be *C-O*. More precisely we have Theorem 4.

Definition. Let $f(x)$ be a continuous function in $a < x < b$. We say that ξ is an odd zero of $f(x)$ in (a,b) if $a < \xi < b$ and

$$\lim_{h \to 0} \operatorname{sgn} f(\xi + h) \operatorname{sgn} f(\xi - h) = -1.$$

THEOREM 4. *Let $\{f_n(x)\}$ $(n = 0, 1, 2, \ldots)$ be a sequence of functions defined in an open interval (a,b). Let E_n, F_n be the sets of odd zeros and zeros of $f_n(x)$, respectively. The functions $f_n(x)$, and the sets E_n, F_n satisfy for $n = 0, 1, 2, \ldots$:*

(1) $f_{n+k}(x) = a_n(x) f_{n+k-1}(x) - b_n(x) f_n(x)$.
(2) *$f_n(x)$ is continuous in (a,b).*
(3) $Z(F_n) \leqslant [n/k] \equiv \mu(n)$.
(4) $\lim_{x \to b^-} \operatorname{sgn} f_n(x) = +1$.
(5) $\lim_{x \to a^+} \operatorname{sgn} f_n(x) = (-1)^{\mu(n)}$.
(6) *$a_n(x)$, $b_n(x)$ are continuous and positive in (a,b).*

Then:

(a) $n > m \geqslant k \Rightarrow \max\{E_n\} > \max\{E_m\}$.
(b) $Z(E_n) = \mu(n)$.
(c) *$E_n, E_{n+1}, \ldots, E_{n+k}$ are C-O for $n = 0, 1, 2, \ldots$.*

Proof. For $0 \leqslant N < k$, $Z(E_n) \leqslant Z(F_n) = 0$. Since F_N is empty, $f_N(x) > 0$ in (a,b). By (2), (4), and (5), $Z(E_k) \geqslant 1$ and hence by (3) $Z(E_k) = 1$. Therefore (a), (b), and (c) are satisfied by the sets $E_0, E_1, \ldots, E_k$. We complete the proof by induction.

We adopt the notation $x(n,j)$ for the elements of E_n (whenever E_n is nonempty) with $x(n,1) > x(n,2) > \ldots$.

Let us assume that for some $N \geqslant k$,

(a*) $x(n,1) > x(m,1)$ for $k \leqslant m < n \leqslant N$,
(b*) $Z(E_n) = \mu(n)$ for $0 \leqslant n \leqslant N$,
(c*) $E_n, E_{n+1}, \ldots, E_{n+k}$ are *C-O*, for $0 \leqslant n \leqslant N - k$.

To complete the proof we must show that:

(a′) $x(N+1, 1) > x(N, 1)$,
(b′) $Z(E_{N+1}) = \mu(N+1)$,
(c′) $E_{N-k+1}, E_{N-k+2}, \ldots, E_{N+1}$ are *C-O*.

From (1) we obtain $\operatorname{sgn} f_{N+1}(x(N,1)) = -\operatorname{sgn} f_{N-k+1}(x(N,1))$. But from (2), (4), (6), and (a*) we have $\operatorname{sgn} f_{N-k+1}(x(N,1)) = +1$. Hence $\operatorname{sgn} f_{N+1}(x(N,1)) = -1$, and by (2) and (4) $x(N+1, 1) > x(N,1)$, which proves (a′).

In the remainder of the proof, m will denote an integer for which $N - k < m \leqslant N$. The symbol $I(j)$ will denote the smallest interval (possibly empty or a single point) which contains $x(m,j)$ for all such m. $I(j)$ is closed. Let j be the smallest integer for which $I(j)$ and $I(j+1)$ are not disjoint. Then there are two distinct values of m, say m' and m'', for which $x(m', j+1) \geqslant x(m'', j)$. But this is impossible since the sets $E_{m'}$, $E_{m''}$ are *C-O* by (c*) and Theorem 2 and if $x(m', j+1) = x(m'', j)$, C1 is violated and if

$x(m', j+1) > x(m'', j)$, C2 is violated in the closed interval $[x(m', j+1), x(m', 1)]$. Hence the intervals $I(j)(j = 1, 2, \ldots)$ are mutually disjoint.

For fixed j, the points $x(m, j)$ are monotonic with respect to m. Specifically $x(N, j) > x(N-1, j) > \ldots$. This is clearly so for $j = 1$, by (a*). For $j > 1$, it is a consequence of Theorem 3 applied to the *C-O* sets $\{E_m\}$.

Let F denote $\bigcup_m E_m$. If $kj \leqslant N$ the set $F \cap I(j)$ has exactly k elements. The set $I(\mu(N+1))$ has $(N+1) - k\mu(N+1)$ elements.

Since by (c*) E_N and E_{N+1-k} are *C-O*, it follows that sgn $f_N(x(N-k+1, j))$ takes the values of ± 1 alternately as j assumes consecutive integer values. This is because all the zeros of $f_N(x)$ are odd zeros and there is precisely one zero of $f_N(x)$ between consecutive zeros of $f_{N-k+1}(x)$. We also know that sgn $f_N(x(N-k+1, 1)) = -1$. Hence we obtain sgn $f_N(x(N-k+1, j) = (-1)^j$. Similarly, sgn $f_{N-k+1}(x(N, j)) = (-1)^{j+1}$. Using (1) we then obtain sgn $f_N(x(N, j) =$ sgn $(x(N-k+1, j)) = (-1)^j$. This means that at the extreme points of $I(j)$, sgn $f_{N+1}(x) = (-1)^j$ for $j = 1, 2, \ldots, \mu(N)$. But the sets are disjoint. Hence between each consecutive pair of non-empty sets $I(j)$, $f_{N+1}(x)$ must have an odd zero. Also there is an odd zero between $I(1)$ and b. Thus $f_{N+1}(x)$ has at least $\mu(N)$ zeros to the right of $I(\mu(N))$. We have already seen that $I(j)$ is non-empty for $j \leqslant \mu(N)$. If $\mu(N) = \mu(N+1)$ then by (3), $Z(E_{N+1}) = \mu(N+1)$. If $\mu(N) \neq \mu(N+1)$, then the signs of $f_{N+1}(x)$ at the extreme points of $I(\mu(N))$ and at $(a+0)$ are different and hence there is another odd zero between a and $I(\mu(N))$. Hence again $Z(E_{N+1}) = \mu(N+1)$. Thus (b') is proved and also since E_{N+1} and $I(j)$ is disjoint for every j, we have that C1 is satisfied by E_m and E_{N+1} for all m in $N-k < m \leqslant N$.

To prove (c') it suffices because of Theorem 2 to show that E_m and E_{N+1} are *C-O*. We have already established C1 for these sets and it remains only to establish C2.

If C2 is not satisfied by E_m and E_{N+1}, there is some interval, J, in which $|Z_J(E_m) - Z_J(E_{N+1})| > 1$. Hence there must be a subinterval J' in which one of $Z_{J'}(E_m)$ and $Z_{J'}(E_{N+1})$ is 0 and the other is 2.

If $Z_{J'}(E_m) = 2$, J' must intersect $I(j)$ for two values of j. But we have seen that between consecutive intervals $I(j)$ there is a point of E_{N+1} and if $Z_{J'}(E_{N+1}) = 2$, J' must contain an interval $I(j)$, since between consecutive intervals $I(j)$ there is precisely one point of E_{N+1} by (3).

Hence C2 is satisfied by E_{N+1}, and E_m for $N-k < m \leqslant N$ and thus (c') is proved.

4. Applications of the principal theorem. Let $\{p_n(x)\}$ denote a sequence of orthonormal polynomials in an interval symmetric about the origin, with respect to an even non-negative weight function. We call such sequences "Symmetric orthonormal polynomials." Then the well-known recurrence formula **(4)** specializes to

$$p_{n+1}(x) = A_n x p_n(x) - B_n p_{n-1}(x) \qquad (n \geqslant 1)$$

where A_n, B_n are positive constants. For $k = 2$, the polynomials $\{p_n(x)\}$ satisfy the hypotheses of Theorem 5 in the interval $(0, \infty)$. To see this we note that (1) is satisfied since they satisfy the recursion formula. Since they are polynomials, they are continuous and (2) is satisfied. Condition 3 is satisfied since the zeros are symmetric about the origin. They satisfy (4) since the normalization which asserts that $A_n > 0$, implies that the coefficient of x^n in $p_n(x)$ is positive. To show that (5) is satisfied we note that for n odd, $\operatorname{sgn} p_{n+1}(0) = -\operatorname{sgn} p_{n-1}(0) \neq 0$ and for n even

$$\operatorname{sgn} \frac{p_{n+1}(x)}{x}\bigg|_{x=0} = -\operatorname{sgn} \frac{p_{n-1}(x)}{x}\bigg|_{x=0}.$$

Hence it is only necessary to verify (5) for $n = 0, 1$. But $p_0(x)$ is a positive constant and $p_1(x) = \alpha x$ where $\alpha > 0$.

Hence Theorem 4 applies and we have the following result:

THEOREM 5. *Let $\{p_n(x)\}$ be a sequence of symmetric orthonormal polynomials. Let P_n denote the set of positive zeros of $p_n(x)$. Then P_n, P_{n+1}, P_{n+2} are C-O for $n = 0, 1, 2, \ldots$.*

Theorem 5 supplements the well-known fact that E_n, E_{n+1} are *C-O* where E_n denotes the set of zeros of $p_n(x)$. We cannot expect E_n, E_{n+1}, E_{n+2} to be *C-O*, since C2 would be violated. Similarly we cannot expect P_n, P_{n+1}, P_{n+2}, P_{n+3} to be *C-O*.

We mention here without proof that Theorem 5 could be proved independently of Theorem 4, by using only the following well-known properties of symmetric orthonormal polynomials **(4)**.

(1) E_n, E_{n+1} are *C-O*.

(2) If $n > m$, there is a point of E_n in any open interval bounded by points of E_m.

(3) E_n is symmetric about zero.

(4) $Z(E_n) = n$.

We now consider another example. Let $p_n(t, k) = p_n = p_n(t)$ be defined by the generating function:

$$F(x,t) = \exp(x + tx^k) = \sum_{n=0}^{\infty} p_n(t;k)x^n. \tag{4.1}$$

As usual, k denotes some integer > 1. Differentiating both sides of (4.1) with respect to x we obtain

$$(1 + kt\,x^{k-1}) \sum_{n=0}^{\infty} p_n x^n = \sum_{n=0}^{\infty} n p_n x^{n-1}\ . \tag{4.2}$$

If we equate coefficients we obtain

$$mp_m = p_{m-1} + ktp_{m-k} : m > k\,. \tag{4.3}$$

We wish to apply Theorem 4 to the sequence $\{p_n(t)\}$ in the interval $(-\infty, 0)$.

Now $\exp(x + tx^k) = \exp(x)\exp(tx^k)$ and hence

$$p_n(t) = \sum_{m+kq=n} \frac{1}{m!}\frac{t^q}{q!} \tag{4.4}$$

where the summation is extended over all non-negative integers m, q satisfying the indicated condition.

This shows that $p_n(t)$ is a polynomial of precise degree $[n/k]$ with all coefficients positive. Hence conditions (2), (3), (4), (5) of Theorem 4 are satisfied.

We use the following consequences of (4.4) in showing conditions (4) and (5) are satisfied.

$$p_n(0) = \frac{1}{n!}$$

$$p_n(t) \sim \frac{t^{[n/k]}}{(n - k[n/k])! \cdot [n/k]!} \text{ as } |t| \to \infty .$$

Hence if E_m denotes the set of zeros of $p_m(t;k)$ which are negative, the sets $E_n, E_{n+1}, \ldots E_{n+k}$ are *C-O*.

When $k = 2$, this result can be interpreted in terms of the Hermite polynomials, $H_n(x)$ defined by

$$\sum_{n=0}^{\infty} \frac{H_n(x)}{n!} w^n = \exp(2xw - w^2). \tag{4.5}$$

Hence

$$H_n(z) = 2^n n! p_n\left(-\frac{1}{4z^2}, 2\right) \cdot z^n$$

and if $y_1, y_2, \ldots$ are zeros of $p_n(y)$ then

$$Z_j = 2\sqrt{1/y_j}$$

are positive zeros of $H_n(x)$. Hence the sets of positive zeros of $H_n(x)$, $H_{n+1}(x)$, $H_{n+2}(x)$ are *C-O*. This was obtained earlier as a special case of the result concerning symmetric orthogonal polynomials.

5. Further examples. We have already seen that the sets of three consecutive Legendre polynomials are *C-O*. For fixed $n \geqslant 2$, let Q, Q', Q'' denote the sets of positive zeros of $P_n(x)$, $P'_n(x)$, $P''_n(x)$ respectively. We shall show that Q, Q', Q'' are *C-O*. The Legendre polynomials satisfy two important functional equations:

$$(n + 1)P_{n+1}(x) - (2n + 1)P_n(x) + nP_{n-1}(x) = 0; \tag{5.1}$$

and

$$(1 - x^2)P''_n(x) - 2xP'_n(x) + n(n + 1)P_n(x) = 0. \tag{5.2}$$

In Theorem 4, we used (5.1) to deduce $\operatorname{sgn} P_{n+1}(x)$ at various critical points. Similarly, we may use (5.2) to deduce $\operatorname{sgn} P_n(x)$ at the zeros of $P'_n(x)$ and $P''_n(x)$.

$$\text{For } x \in Q, \quad \operatorname{sgn} P'_n(x) = \operatorname{sgn} P''_n(x).$$

$$\text{For } x \in Q', \quad \operatorname{sgn} P_n(x) = -\operatorname{sgn} P''_n(x).$$

$$\text{For } x \in Q'', \quad \operatorname{sgn} P_n(x) = \operatorname{sgn} P'_n(x).$$

Also since $Q \subset (0,1)$, we have by the standard normalization,

$$\operatorname{sgn} P_n(1) = \operatorname{sgn} P'_n(x) = \operatorname{sgn} P''_n(1) = 1.$$

From (5.2) and the simplicity of the zeros of $P_n(x)$ it is clear that Q and Q'' are disjoint. It remains only to verify C2.

Let us assume an interval J for which $Z_J(Q) > Z_J(Q'') + 1$. Then there is an interval $I \subset J$, such that

$$Z_I(Q) = 2, \qquad Z_I(Q'') = 0.$$

Since Q, Q' are *C-O*, there is precisely one point of Q' in I. Let x_1, x_2 be the two points of Q in I. Then $\operatorname{sgn} P'_n(x_1) = -\operatorname{sgn} P'_n(x_2)$, and hence $\operatorname{sgn} P''_n(x_1) = -\operatorname{sgn} P''_n(x_2)$, and thus $Z_I(Q'') \geqslant 1$. But this is a contradiction. Hence there is no interval J in which $Z_J(Q) > Z_J(Q'') + 1$. Similarly there is no interval J in which $Z_J(Q'') > Z_J(Q) + 1$. Hence C2 is satisfied and Q, Q'' are *C-O*.

References

1. Maxime Bôcher, *On certain methods of Sturm and their applications to the roots of Bessel's functions*, Bull. Amer. Math. Soc., *3* (1897), 205–213.
2. Albert Edrei, *Proof of a conjecture of Schoenberg on the generating function of a totally positive sequence*, Can. J. Math., *5* (1953), 86–94.
3. I. J. Schoenberg, *Some analytic aspects of the problem of smoothing*, Studies and Essays Presented to R. Courant on his 60th Birthday (New York, 1948), 351–369.
4. Gabor Szegö, *Orthogonal polynomials*, Amer. Math. Soc. Coll. Publ. *23* (New York, 1939), 41.

Johns Hopkins University

CERTAIN GROUPS OF ORTHONORMAL STEP FUNCTIONS

J. J. PRICE

1. Introduction. It was first pointed out by Fine **(2)**, that the Walsh functions are essentially the characters of a certain compact abelian group, namely the countable direct product of groups of order two. Later Chrestenson **(1)** considered characters of the direct product of cyclic groups of order α ($\alpha = 2, 3, \ldots$). In general, his results show that the analytic properties of these generalized Walsh functions are basically the same as those of the ordinary Walsh functions. That this is the case does not seem surprising, since the structures of the underlying groups are quite similar.

It is our purpose to show that a considerably different state of affairs prevails for characters of a direct product of finite cyclic groups whose orders are unbounded. We shall show by examples that for these functions, certain basic properties of the Walsh systems mentioned above no longer hold.

The functions we shall study as well as the Walsh systems are included in a very general class of step functions discussed recently by Ohkuma **(3)**. We shall quote some of his results. However, a number of his principal theorems are proved under special conditions which are not satisfied by the functions we shall consider.

We shall have occasion to cite a number of properties of the Walsh functions. Unless otherwise stated, the reference for all of these is the fundamental paper of Fine **(2)**.

2. Definitions. Let $\{n_1, n_2, \ldots\}$ be a sequence of integers, $n_i \geqslant 2$, and let $p_0 = 1, p_k = n_1 n_2 \ldots n_k$. Denote by $I(r, k)$ the interval $r/p_k \leqslant x < (r+1)/p_k$ and by $\chi_r^{(k)}(x)$ the characteristic function of the set $\bigcup I(j, k)$ where j runs over all integers $\equiv r \pmod{p_k}$. Let us define a system of functions $\{\phi_0(x), \phi_1(x), \ldots\}$ as follows.

$$\phi_{k-1}(x) = \sum_{r=0}^{p_k - 1} \omega_k{}^r \chi_r^{(k)}(x), \qquad \omega_k = e^{2\pi i/n_k} \qquad (k = 1,2, \ldots). \tag{1}$$

By definition, $\phi_{k-1}(x)$ is a step function whose values run through the n_kth roots of unity. On the interval [0,1] there are p_k intervals of constancy each of length $1/p_k$.

Now let $\Psi(n_1, n_2, \ldots)$ be the set of all finite products of these functions. In particular, $\Psi(2,2,2, \ldots)$ is the system of Walsh functions and $\Psi(\alpha, \alpha, \alpha, \ldots)$ is the generalization of Chrestenson.

Received June 8, 1956. This paper is essentially half of a doctoral thesis at the University of Pennsylvania. The author gratefully acknowledges the sponsorship of Dr. N. J. Fine.

To enumerate our system we use a scheme based on Paley's enumeration of the Walsh functions. Every positive integer N can be expressed uniquely in the form $r_0p_0 + r_1p_1 + \ldots + r_mp_m$ where $0 \leqslant r_i < n_{i+1}$. Set

$$\psi_0(x) \equiv 1, \qquad \psi_N(x) = \phi_0^{r_0}(x)\phi_1^{r_1}(x) \ldots \phi_m^{r_m}(x). \tag{2}$$

It is clear that $\Psi(n_1, n_2, \ldots)$ is an orthonormal system. That it is complete can be seen in several ways. For instance, the reasoning used by Fine in **(2)** shows that it is the full set of characters of the countable direct product of cyclic groups of order n_i transferred to the unit interval in a measure-preserving manner. Or, one may invoke a general theorem of Ohkuma **(3)**.

3. Expansions of certain functions. In this section, we obtain the Fourier expansions of several functions with respect to the system $\Psi(n_1, n_2, \ldots)$.

THEOREM 1.

$$x - [x] \sim \tfrac{1}{2} + \sum_{k=1}^{\infty} \frac{1}{p_k} \sum_{j=1}^{n_k-1} \frac{1}{\omega_k^{-j} - 1}\, \phi_{k-1}^{\,j}(x).$$

Proof. For every real value of x, $x - [x]$ can be represented in the form

$$\sum_{k=1}^{\infty} \frac{c_k}{p_k}, \qquad 0 \leqslant c_k < n_k .$$

This representation is unique except when x is of the form m/p_k. Now let $\zeta_\nu^{(k)}(x)$ be the characteristic function of the set of all x for which $\phi_{k-1}(x) = \omega_k^{\nu}$. This set can also be described as $\bigcup I(r,k)$ where the union is taken over all integers $r \equiv \nu \pmod{n_k}$. Define

$$c_k(x) = \sum_{\nu=0}^{n_k-1} \nu \zeta_\nu^{(k)}(x) . \tag{3}$$

Then, for every value of x,

$$x - [x] = \sum_{k=1}^{\infty} \frac{c_k(x)}{p_k} . \tag{4}$$

In the case that $x = m/p_k$, (4) gives the finite representation of $x - [x]$. In all cases, the series in (4) converges uniformly. We observe that

$$\frac{1}{n_k} \sum_{j=0}^{n_k-1} \omega_k^{-j\nu}\, \phi_{k-1}^{j}(x) = \begin{cases} 1, \text{ if } \phi_{k-1}(x) = \omega_k^{\nu} \\ 0, \text{ otherwise.} \end{cases} \tag{5}$$

Thus, the series (5) is the Fourier expansion of $\zeta_\nu^{(k)}(x)$. Substituting it in (3) yields

$$c_k(x) = \frac{1}{n_k} \sum_{j=0}^{n_k-1} \phi_{k-1}^{\,j}(x) \sum_{\nu=0}^{n_k-1} \nu(\omega_k^{-j})^{\nu} . \tag{6}$$

Now in general,

$$\sum_{\nu=0}^{N} \nu x^{\nu} = x \frac{d}{dx}\left(\frac{1 - x^{N+1}}{1 - x}\right) = x \frac{Nx^{N+1} - (N+1)x^N + 1}{(1-x)^2} .$$

Putting $N = n_k - 1$, $x = \omega_k^{-j}$, and using the fact that $\omega_k^{n_k} = 1$,

$$\sum_{\nu=0}^{n_k-1} \nu(\omega_k^{-j})^\nu = \frac{n_k(\omega_k^{-j} - 1)}{(1 - \omega_k^{-j})^2} = \frac{n_k}{\omega_k^{-j} - 1} \qquad (j \neq 0), \tag{7}$$

$$\sum_{\nu=0}^{n_k-1} \nu(\omega_k^{-j})^\nu = \sum_{\nu=1}^{n_k-1} \nu = \frac{n_k(n_k - 1)}{2} \qquad (j = 0). \tag{8}$$

Substituting (7) and (8) in (6),

$$c_k(x) = \frac{n_k - 1}{2} + \sum_{j=1}^{n_k-1} \frac{1}{\omega_k^{-j} - 1} \phi_{k-1}{}^j(x). \tag{9}$$

Finally, using (9), the series in (4) becomes

$$\tfrac{1}{2} \sum_{k=1}^{\infty} \frac{n_k - 1}{p_k} + \sum_{k=1}^{\infty} \frac{1}{p_k} \sum_{j=1}^{n_k-1} \frac{1}{\omega_k^{-j} - 1} \phi_{k-1}{}^j(x). \tag{10}$$

Since the convergence in (4) is uniform, it follows from (9) that there is a sequence of partial sums of the series (10) converging uniformly to $x - [x]$. This implies that (10) is the Fourier series of $x - [x]$.

If we observe that

$$\sum_{k=1}^{\infty} \frac{n_k - 1}{p_k} = \sum_{k=1}^{\infty} \left(\frac{1}{p_{k-1}} - \frac{1}{p_k} \right) = \frac{1}{p_0} - \lim_{k \to \infty} \frac{1}{p_k} = \frac{1}{p_0} = 1,$$

then (10) is the series given in the statement of the theorem.

The next theorem involves a simple, but rather lengthy computation. For this reason, we introduce the following lemma which prepares a computation needed in the proof of the theorem. Define

$$s(\nu, k) = 1 + \omega_k + \omega_k{}^2 + \ldots + \omega_k{}^{\nu-1} - \nu\omega_k{}^\nu.$$

LEMMA 1.

$$\sum_{\nu=0}^{p_k-1} \omega_k{}^{-\nu m} s(\nu, k) = \begin{cases} \dfrac{2p_k}{1 - \omega_k} & (m = 0), \\ -\dfrac{p_k}{1 - \omega_k} - \dfrac{p_k(p_k - 1)}{2} & (m = 1), \\ \dfrac{p_k}{1 - \omega_k{}^{1-m}} & (m = 2,3, \ldots, n_k - 1). \end{cases}$$

Proof.

$$\sum_{\nu=0}^{p_k-1} \omega_k{}^{-\nu m}(1 + \omega_k + \omega_k{}^2 + \ldots + \omega_k{}^{\nu-1}) = \sum_{\nu=0}^{p_k-1} \omega_k{}^{-\nu m} \frac{1 - \omega_k{}^\nu}{1 - \omega_k} \tag{11}$$

$$= \frac{1}{1 - \omega_k} \left[\sum_{\nu=0}^{p_k-1} \omega_k{}^{-\nu m} - \sum_{\nu=0}^{p_k-1} \omega_k{}^{\nu(1-m)} \right] = \begin{cases} \dfrac{p_k}{1 - \omega_k} & (m = 0), \\ -\dfrac{p_k}{1 - \omega_k} & (m = 1), \\ 0 & (m = 2,3, \ldots, n_k - 1). \end{cases}$$

Applying formulas (7) and (8), with n_k replaced by p_k,

$$(12)\quad -\sum_{\nu=0}^{p_k-1} \omega_k^{-\nu m}(\nu\omega_k^{\nu}) = -\sum_{\nu=0}^{p_k-1} \nu(\omega_k^{1-m})^{\nu} = \begin{cases} -\dfrac{p_k(p_k-1)}{2} & (m=1), \\ \dfrac{p_k}{1-\omega_k^{1-m}} & (m=0,2,3,\ldots,n_k-1). \end{cases}$$

Adding (11) and (12) yields the lemma.

We shall now find the Fourier series associated with the function

$$J_{k-1}(x) = \int_0^1 \phi_{k-1}(u)du.$$

THEOREM 2.

$$J_{k-1}(x) \sim \frac{1}{p_k(1-\omega_k)} + \phi_{k-1}(x)\left[\frac{1}{2p_k} - \frac{1}{p_k(1-\omega_k)}\right] + \sum_{\nu=k+1}^{\infty} \frac{1}{p_\nu} \sum_{j=1}^{n_\nu-1} \frac{1}{\omega_\nu^{-j}-1} \psi_{jp_{\nu-1}+p_{k-1}}(x).$$

Proof. If $x \in I(\nu, k)$,

$$\phi_{k-1}(x) = \omega_k^{\nu}; \int_0^x \phi_{k-1}(u)du = \int_0^{\nu/p_k} \phi_{k-1}(u)du + \omega_k^{\nu}\left(x - [x] - \frac{\nu}{p_k}\right).$$

Therefore, if $x \in I(\nu, k)$,

$$J_{k-1}(x) = \omega_k^{\nu}\left(x - [x] - \frac{\nu}{p_k}\right) + \frac{1}{p_k}\sum_{j=0}^{\nu-1} \omega_k^{j} = \omega_k^{\nu}(x - [x]) + \frac{1}{p_k} s(\nu,k).$$

Since for all x,

$$J_{k-1}(x) = \sum_{\nu=0}^{p_k-1} \chi_\nu^{(k)}(x)\, J_{k-1}(x),$$

$$(13)\quad J_{k-1}(x) = (x - [x])\sum_{\nu=0}^{p_k-1} \omega_k^{\nu}\chi_\nu^{(k)}(x) + \frac{1}{p_k}\sum_{\nu=0}^{p_k-1} s(\nu,k)\chi_\nu^{(k)}(x).$$

The first sum on the right side of (13) is $\phi_{k-1}(x)$ by definition.

Furthermore,

$$\chi_\nu^{(k)}(x) = \frac{1}{p_k}\sum_{j=0}^{p_k-1} \overline{\psi_j(\nu,k)}\, \psi_j(x),$$

where $\psi_j(\nu, k)$ is the value of $\psi_j(x)$ on $I(\nu, k)$. Substituting these facts into (13),

$$(14)\quad J_{k-1}(x) = (x - [x])\,\phi_{k-1}(x) + \frac{1}{p_k^2}\sum_{j=0}^{p_k-1} \psi_j(x)\sum_{\nu=0}^{p_k-1} \overline{\psi_j(\nu,k)}\, s(\nu,k).$$

We assert now that the Fourier expansion of $J_{k-1}(x)$ does not involve any of the functions $\psi_j(x)$ for which $1 \leqslant j < p_{k-1}$. To show that the corres-

ponding Fourier coefficients vanish, we use the facts that $J_{k-1}(x)$ has period $1/p_{k-1}$, $\psi_j(x)$ is constant on intervals of the form $I(r, k-1)$, and

$$\int_0^1 \psi_j(x)dx = 0.$$

Consequently,

$$\int_0^1 J_{k-1}(x)\,\overline{\psi_j(x)}\,dx = \int_0^{1/p_{k-1}} J_{k-1}(x)dx \int_0^1 \overline{\psi_j(x)}dx = 0.$$

Now if $p_{k-1} \leqslant j < p_k$, then $\psi_j(x)$ can occur in the expansion of $J_{k-1}(x)$ only if $\psi_j(x)$ is one of the functions

$$\phi_{k-1}^m(x), \qquad (m = 1, 2, \ldots, n_k - 1).$$

If $\psi_j(x)$ is not of this type, it is of the form $\phi_{k-1}{}^m(x)\psi_r(x)$, where $1 \leqslant r < p_{k-1}$. But then $J_{k-1}(x)\phi_{k-1}{}^{-m}(x)$ has period $1/p_{k-1}$ and the same reasoning shows that

$$\int_0^1 J_{k-1}(x)\,\overline{\psi_j(x)}\,dx = \int_0^1 (J_{k-1}(x)\,\phi_{k-1}{}^{-m}(x))\,\overline{\psi_r(x)}dx = 0.$$

The expansion of $x - [x]$ is given by Theorem 1. We substitute it into (14). In view of the above discussion, however, we may drop all terms in $\psi_j(x)$ for $1 \leqslant j < p_k$ except those of the type $\phi_{k-1}{}^m(x)$. The result is:

$$\begin{aligned}(15)\qquad J_{k-1}(x) \sim {} & \tfrac{1}{2}\phi_{k-1}(x) + \frac{1}{p_k}\sum_{j=1}^{n_k-1} \frac{1}{\omega_k^{-j} - 1}\,\phi_{k-1}{}^{j+1}(x) \\ & + \sum_{r=k+1}^{\infty} \frac{1}{p_r}\sum_{j=1}^{n_r-1} \frac{1}{\omega_r^{-j} - 1}\,\phi_{r-1}{}^j(x)\,\phi_{k-1}(x) \\ & + \frac{1}{p_k^2}\sum_{m=0}^{n_k-1} \phi_{k-1}{}^m(x)\sum_{\nu=0}^{p_k-1} \phi_{k-1}{}^{-m}(\nu,k)\,s(\nu,k).\end{aligned}$$

The first sum on the right side of (15) has a term in $\phi_{k-1}{}^{n_k}(x) \equiv 1$. Writing this term first and shifting the summation index,

$$(16)\qquad \frac{1}{p_k}\sum_{j=1}^{n_k-1} \frac{1}{\omega_k^{-j} - 1}\,\phi_{k-1}{}^{j+1}(x) = -\frac{1}{p_k(1-\omega_k)} - \frac{1}{p_k}\sum_{m=2}^{n_k-1} \frac{1}{1-\omega_k{}^{1-m}}\,\phi_{k-1}^m(x).$$

Since $\phi_{k-1}{}^{-m}(\nu, k) = \omega_k{}^{-\nu m}$, the last sum in (15) can be simplified by means of Lemma 1. We obtain

$$\begin{aligned}(17)\qquad & \frac{1}{p_k^2}\sum_{m=0}^{n_k-1} \phi_{k-1}{}^m(x)\sum_{\nu=0}^{p_k-1} \omega_k{}^{-\nu m}s(\nu,k) \\ & = \frac{2}{p_k(1-\omega_k)} - \phi_{k-1}(x)\left[\frac{1}{p_k(1-\omega_k)} + \tfrac{1}{2} - \frac{1}{2p_k}\right] \\ & \qquad + \frac{1}{p_k}\sum_{m=2}^{n_k-1} \frac{1}{1-\omega_k{}^{1-m}}\,\phi_{k-1}^m(x).\end{aligned}$$

Substituting (16) and (17) into (15),

$$(18)\qquad J_{k-1}(x) \sim \frac{1}{p_k(1-\omega_k)} + \phi_{k-1}(x)\left[\frac{1}{2p_k} - \frac{1}{p_k(1-\omega_k)}\right] + \sum_{\nu=k+1}^{\infty} \frac{1}{p_\nu} \sum_{j=1}^{n_\nu - 1} \frac{1}{\omega_\nu^{-j} - 1}\, \phi_{\nu-1}{}^{j}(x)\, \phi_{k-1}(x).$$

In our enumeration,

$$\phi_{\nu-1}{}^{j}(x)\, \phi_{k-1}(x) = \psi_{jp_{\nu-1}+p_{k-1}}(x).$$

Therefore, (18) is the desired expansion.

It is easy to obtain a similar expansion for the integral of the function $\phi_{k-1}{}^{m}(x)$. Basically, this amounts to replacing ω_k by $\omega_k{}^{m}$ in (18). We can now obtain the Fourier series associated with the integral of the function $\psi_j(x)$ for any value of j.

Given j, there exists k such that $p_{k-1} \leqslant j < p_k$. Then $j = mp_{k-1} + r$ where $1 \leqslant m \leqslant n_k - 1$ and $0 \leqslant r < p_{k-1}$ so that $\psi_j(x) = \phi_{k-1}{}^{m}(x)\psi_r(x)$. Now the integral of $\phi_{k-1}{}^{m}(x)$ has period $1/p_{k-1}$ whereas the period of $\psi_r(x)$is a multiple of this number. Therefore,

$$\int_0^x \psi_j(u)du = \psi_r(x) \int_0^x \phi_{k-1}{}^{m}(u)du.$$

and we easily obtain the following expansion.

THEOREM 3. *Let* $j = mp_{k-1} + r$ *where* $1 \leqslant m \leqslant n_{k-1}$ *and* $0 \leqslant r < p_{k-1}$. *Then*

$$\int_0^x \psi_j(u)du \sim \frac{1}{p_k(1-\omega_k^{m})}\psi_r(x) + \left[\frac{1}{2p_k} - \frac{1}{p_k(1-\omega_k^{m})}\right]\psi_j(x) + \sum_{\nu=k+1}^{\infty} \frac{1}{p_\nu} \sum_{t=1}^{n_\nu - 1} \frac{1}{\omega_\nu^{-t} - 1}\, \psi_{tp_{\nu-1}+j}(x).$$

4. Fourier coefficients. The following theorems are of interest.

(A) If $f(x)$ is a continuous function with modulus of continuity $\omega(\delta)$ and if $p_{k-1} \leqslant j < p_k$, then $|a_j| \leqslant \frac{1}{2}\omega(1/p_{k-1})$ where a_j is the jth Fourier coefficient of $f(x)$ with respect to the system $\Psi(n_1, n_2, \ldots)$.

(B) If $f(x)$ has bounded total variation V, then

$$|a_j| \leqslant \frac{V}{p_k} \csc \pi/n_k$$

Theorem (A) is a particular case of a result of Ohkuma **(3)**; Theorem (B) is essentially a result of Chrestenson **(1)**. When the sequence $\{n_1, n_2, \ldots\}$ is bounded, there follow directly from (A) and (B) such facts as:

(a) If $f(x)$ satisfies a Lipschitz condition of order α, then its Fourier coefficients are $O(1/n^{\alpha})$.

(b) If $f(x)$ is of bounded variation, its coefficients are $O(1/n)$.

However, if the sequence is unbounded, these assertions cannot be made. For instance, if $f(x)$ satisfies a Lipschitz condition of order 1, the most we

can say about a_{p_k-1} is that it is less in absolute value than C/p_{k-1} where C is a constant. Consequently we can conclude only that $|a_j| < Cn_k/j$ when $p_{k-1} \leqslant j < p_k$. Since $\limsup n_k = \infty$, it is conceivable that $\limsup j|a_j| = \infty$. Furthermore, $\limsup V \csc \pi/n_k = \infty$, so it is also conceivable that $\limsup j|a_j| = \infty$ for functions of bounded variation. We shall show by means of examples that these possibilities actually do occur.

THEOREM 4. *If $\{n_1, n_2, \ldots\}$ is unbounded, then there exist functions of bounded variation whose Fourier coefficients with respect to the system $\Psi(n_1, n_2, \ldots)$ are not $O(1/n)$.*

Proof. Take the function $f(x) = x - [x]$ whose expansion is given by Theorem 1. Let a_j denote the jth Fourier coefficient of $f(x)$. For any k, $\phi_{k-1}{}^{n_k-1}(x)$ is the $(n_k - 1)p_{k-1}$th or the $(p_k - p_{k-1})$th function in our enumeration of the system Ψ. By Theorem 1,

$$a_{p_k-p_{k-1}} = \frac{1}{p_k(\omega_k - 1)} = \frac{1}{p_k} \frac{1}{e^{2\pi i/n_k} - 1} .$$

$$(p_k - p_{k-1}) \mid a_{p_k-p_{k-1}} \mid = \frac{p_k - p_{k-1}}{2p_k} \csc \pi/n_k .$$

$$\limsup (p_k - p_{k-1}) \mid a_{p_k-p_{k-1}} \mid = \limsup \tfrac{1}{2}\left(1 - \frac{1}{n_k}\right) \csc \pi/n_k = \infty .$$

Hence, $\limsup j|a_j| = \infty$.

THEOREM 5. *If $\{n_1, n_2, \ldots\}$ is unbounded there exist absolutely continuous functions, in fact functions satisfying a Lipschitz condition of order 1, whose Fourier coefficients with respect to the system $\Psi(n_1, n_2, \ldots)$ are not $O(1/n)$.*

Proof. Take the function $J_{k-1}(x)$. Formula (14) in the proof of Theorem 2 shows that except for its first p_k terms, the Fourier series of $J_{k-1}(x)$ agrees with that of $(x - [x])\phi_{k-1}(x)$. Except for a shift, the Fourier coefficients of the latter function are the same as those of $x - [x]$. Consequently the argument used in the proof of Theorem 4 again shows that $\limsup j|a_j| = \infty$.

The following elementary facts will be useful. We list them as a lemma and sketch the proof.

LEMMA 2. *If $n_1 < n_2 < n_3 < \ldots$, then as $k \to \infty$,*

$$\frac{1}{n_k} \sum_{j=1}^{n_k-1} \frac{1}{|\omega_k{}^{-j} - 1|} \sim \frac{1}{n_k} \left| \sum_{j=1}^{[n_k/4]} \frac{1}{\omega_k{}^{-j} - 1} \right| \sim \log n_k .$$

Proof. Since $|\omega_k{}^{-j} - 1|^{-1} = |e^{-2\pi ij/n_k} - 1|^{-1} = \frac{1}{2} \csc \pi j/n_k$, the first sum is clearly asymptotic to

$$\frac{1}{n_k} \sum_{j=1}^{[n_k/2]} \csc \pi j/n_k .$$

Here, $\csc \pi j/n_k$ may be replaced by $n_k/\pi j$, since the ratio of $\csc x$ to $1/x$ is bounded for $0 < x \leq \frac{1}{2}\pi$. Therefore, the first sum is asymptotic to

$$\frac{1}{n_k}\sum_{j=1}^{[n_k/2]}\frac{n_k}{\pi j} \sim \log [n_k/2] \sim \log n_k .$$

To prove the second assertion, observe that

$$\frac{1}{n_k}\left|\sum_{j=1}^{[n_k/4]}\frac{1}{\omega_k^{-j}-1}\right| = \frac{1}{n_k}\left|\tfrac{1}{2}\sum_{j=1}^{[n_k/4]}(-1+i\cot \pi j/n_k)\right| \sim \frac{1}{n_k}\sum_{j=1}^{[n_k/4]}\cot \pi j/n_k .$$

Arguing as above, $\cot \pi j/n_k$ may be replaced by $n_k/\pi j$ and the result follows in the same way.

Fine has proved that the Walsh-Fourier coefficients of an absolutely continuous function are not $O(1/n)$ unless the function is a constant. We shall modify his argument to show that there is a class of unbounded sequences $\{n_i\}$ such that the Fourier coefficients of a non-constant absolutely continuous function with respect to $\Psi(n_1, n_2, \ldots)$ are not even $O(1/n)$.

We shall consider sequences $\{n_i\}$ satisfying:

$$n_1 < n_2 < n_3 < \ldots ; \sum_{\nu=k+1}^{\infty}\frac{1}{p_\nu}\sum_{s=1}^{n_\nu - 1}\frac{1}{|\omega_\nu^{-s}-1|} = O(1/p_{k-1}) . \tag{19}$$

In view of Lemma 2, the latter condition is equivalent to

$$\sum_{\nu=k+1}^{\infty}\frac{\log n_\nu}{p_{\nu-1}} = O(1/p_{k-1}) .$$

This is a restriction on the rate of increase of the sequence. Still, it does admit sequences growing as rapidly as

$$2, 2^2, 2^{2^2}, \ldots ; n_{k+1} = 2^{n_k} .$$

THEOREM 6. *If $\{n_1, n_2, \ldots\}$ satisfies* (19) *then the Fourier coefficients of a non-constant absolutely continuous function with respect to the system $\Psi(n_1, n_2, \ldots)$ are not $O(1/n)$.*

Proof. Suppose that $F(x)$ is absolutely continuous and that $f(x)$ is its derivative. Let $\{a_j\}$ and $\{b_j\}$ be the Fourier coefficients of $F(x)$ and $f(x)$ respectively. We assume $F(x)$ is not a constant so that $f(x) \not\equiv 0$. Therefore one of the Fourier coefficients of $f(x)$, say b_r, does not vanish. We shall show that if $i_k = p_k - p_{k-1} + r$ then

$$i_k|a_{i_k}| \backsim n_k .$$

Hence, $\limsup j|a_j| = \infty$.

Set

$$H_j(x) = \int_0^x \psi_j(u)du .$$

Since $H_j(0) = H_j(1) = 0$, we obtain, on integrating by parts,

$$(20)\qquad a_{i_k} = \int_0^1 F(x)\,\overline{\psi_{i_k}(x)}dx = \left[F(x)\,\overline{H_{i_k}(x)}\right]_0^1 - \int_0^1 f(x)\,\overline{H_{i_k}(x)}dx$$
$$= -\int_0^1 f(x)\,\overline{H_{i_k}(x)}dx.$$

Since the Fourier series given by Theorem 3 has a subsequence of partial sums converging uniformly to

$$H_{i_k}(x),$$

we may substitute it in (20) and integrate termwise.

$$-a_{i_k} = \frac{1}{p_k(1-\omega_k)}\int_0^1 f(x)\overline{\psi_r(x)}dx + \left[\frac{1}{2p_k} - \frac{1}{p_k(1-\omega_k)}\right]\int_0^1 f(x)\overline{\psi_{i_k}(x)}dx$$
$$+\sum_{\nu=k+1}^{\infty}\frac{1}{p_\nu}\sum_{s=1}^{n_\nu-1}\frac{1}{\omega_\nu^s-1}\int_0^1 f(x)\overline{\psi_{sp_\nu-1+i_k}(x)}dx.$$

$$(21)\qquad -a_{i_k} = \frac{1}{p_k(1-\omega_k)}b_r + \left[\frac{1}{2p_k} - \frac{1}{p_k(1-\omega_k)}\right]b_{i_k}$$
$$+\sum_{\nu=k+1}^{\infty}\frac{1}{p_\nu}\sum_{s=1}^{n_\nu-1}\frac{1}{\omega_\nu^s-1}b_{sp_\nu-1+i_k}.$$

Now

$$\max_{j>p_k}|b_j| \to 0, \qquad k\to\infty.$$

Therefore, by assumption (19), the absolute value of the sum in (21) is $O(1/p_{k-1}) = O(n_k/p_k)$. Multiplying equation (21) by $-i_k$,

$$(22)\qquad i_k a_{i_k} = -\frac{i_k}{p_k}\frac{b_r}{1-\omega_k} - \frac{i_k}{p_k}\left(\frac{1}{2} - \frac{1}{1-\omega_k}\right)b_{i_k} + o(i_k n_k/p_k).$$

Since

$$\lim_{k\to\infty}\frac{i_k}{p_k} = \lim_{k\to\infty}\frac{p_k - p_{k-1} + r}{p_k} = 1; \frac{1}{1-\omega_k}\sim n_k,$$

the absolute value of the first term in (22) is asymptotic to n_k. The second term is $o(n_k)$ since $b_{i_k}\to 0$. Hence,

$$i_k\,|a_{i_k}| \sim n_k, \qquad k\to\infty.$$

5. Summability. The trigonometric Fourier series of a continuous function $f(x)$ is uniformly $(C,1)$ summable to $f(x)$. The analogous statement is true for the Walsh functions. In fact, a theorem of Ohkuma **(3)** implies it is true in any system $\Psi(n_1, n_2, \ldots)$ provided $\{n_i\}$ is a bounded sequence. We shall show that if $\{n_i\}$ is unbounded, the situation is quite different.

THEOREM 7. *Let $\{n_1, n_2, \ldots\}$ be unbounded. Then, given any real value a, there exists a continuous function whose Fourier series with respect to the system $\Psi(n_1, n_2, \ldots)$ is not $(C,1)$ summable at $x = a$.*

Proof. Denote by $D_k(x, u)$ the kth Dirichlet kernel of the system.

$$D_k(x,u) = \sum_{j=0}^{k-1} \psi_j(x)\overline{\psi_j(u)}. \tag{23}$$

It will be useful to note that if $x \in I(r, k)$, then

$$D_{p_k}(x,u) = \prod_{\nu=0}^{k-1} \left\{ \sum_{j=0}^{n_{\nu+1}-1} (\phi_\nu(x)\phi_\nu^{-1}(u))^j \right\} = p_k \chi_r^{(k)}(u). \tag{24}$$

Denote by $K_n(x, u)$ the nth Cesaro kernel:

$$K_n(x, u) = \frac{1}{n}\sum_{j=1}^{n} D_j(x, u). \tag{25}$$

If the sequence

$$\left\{ \int_0^1 |K_n(a, u)|du \right\}$$

is unbounded, then by a well-known theorem of Haar, there exists a continuous function whose Fourier series is not $(C, 1)$ summable at $x = a$.

Let us first consider the case when $a = 0$. This is a convenient value of a because $\psi_j(0) = 1$ for all j. For brevity, let

$$D_n(u) = D_n(0, u),\ K_n(u) = K_n(0, u),\ K_n = \int_0^1 |K_n(u)|du.$$

We shall show that the sequence $\{K_n\}$ is unbounded.

$$\begin{aligned} p_{r+1}K_{p_{r+1}}(u) &= \sum_{j=1}^{p_{r+1}} D_j(u) = \sum_{j=1}^{p_r} D_j(u) + \sum_{p_r+1}^{p_{r+1}} D_j(u) \\ &= p_r K_{p_r}(u) + \sum_{\nu=1}^{n_{r+1}-1}\sum_{j=1}^{p_r} D_{\nu p_r+j}(u). \end{aligned} \tag{26}$$

Our enumeration of $\Psi(n_1, n_2, \ldots\}$ is such that

$$\begin{aligned} \sum_{j=1}^{p_r} D_{\nu p_r+j}(u) &= \sum_{j=1}^{p_r} \{(1 + \phi_r^{-1}(u) + \phi_r^{-2}(u) + \ldots + \phi_r^{-(\nu-1)}(u))\, D_{p_r}(u) \\ &\qquad + \phi_r^{-\nu}(u)\, D_j(u)\} \\ &= p_r D_{p_r}(u)\,((1 + \phi_r^{-1}(u) + \ldots + \phi_r^{-(\nu-1)}(u)) + \phi_r^{-\nu}(u)\, p_r K_{p_r}(u). \end{aligned} \tag{27}$$

Combining (26) and (27),

$$\begin{aligned} p_{r+1}K_{p_{r+1}}(u) &= p_r D_{p_r}(u)\sum_{\nu=1}^{n_{r+1}-1}(1 + \phi_r^{-1}(u) + \ldots + \phi_r^{-(\nu-1)}(u)) \\ &\qquad + p_r K_{p_r}(u)\sum_{\nu=0}^{n_{r+1}-1}\phi_r^{-\nu}(u). \end{aligned} \tag{28}$$

An easy computation yields

$$K_{p_{r+1}}(u) = \begin{cases} \dfrac{1}{1 - \phi_r^{-1}(u)} D_{p_r}(u) & (\phi_r(u) \neq 1), \\ K_{p_r}(u) + \dfrac{n_{r+1} - 1}{2} D_{p_r}(u) & (\phi_r(u) = 1). \end{cases} \tag{29}$$

By means of (24) and the recursion relation (29), it is possible to obtain an exact formula for $K_{p_r}(u)$. The reasoning involves a number of cases, all of which are quite simple. We shall simply list the results.

$$(30)\quad K_{p_r}(u) = \begin{cases} \dfrac{p_r + 1}{2}, & u \in I(0, r); \\ \dfrac{p_{s-1}}{1 - \omega_s^{-j}}, & u \in I(j\,p_r/p_s, r) \qquad (j = 1, 2, \ldots, n_s - 1; \\ 0, & \text{otherwise.} \qquad\qquad s = 1, 2, \ldots, r); \end{cases}$$

It follows from (30) that

$$(31)\quad \int_0^1 |K_{p_r}(u)|\,d\,u = \frac{1}{p_r}\left\{\frac{p_r + 1}{2} + \sum_{s=1}^{r}\sum_{j=1}^{n_s-1}\frac{p_{s-1}}{|1 - \omega_s^{-j}|}\right\} > \frac{1}{p_r}\sum_{j=1}^{n_r-1}\frac{p_{r-1}}{|1 - \omega_r^{-j}|}$$
$$= \frac{1}{n_r}\sum_{j=1}^{n_r-1}\frac{1}{|1 - \omega_r^{-j}|}.$$

Therefore, by Lemma 2, the sequence $\{K_{p_r}\}$ is unbounded, increasing at least as rapidly as $\{C \log n_r\}$ where C is a constant. Consequently there exists a continuous function whose Fourier series is not $(C, 1)$ summable at the point $x = 0$.

The functions

$$\int_0^1 |K_r(x, u)|du$$

are actually independent of x, hence constant. This is because $\Psi(n_1, n_2, \ldots)$ is essentially the set of characters of a compact group whose Haar measure coincides with the Lebesgue measure on the unit interval. (The argument given by Fine for the Walsh functions (2) carries over directly.) It follows that our proof is valid not only for $a = 0$, but for any value of a. This proves the theorem.

We show next that the situation with respect to summability can be even worse if the sequence $\{n_i\}$ increases rapidly enough. Such a sequence will be one such that

$$(32)\qquad n_1 < n_2 < n_3 < \ldots;\ \frac{1}{p_k}\sum_{r=1}^{[n_k/4]}\frac{1}{\omega_k^{-r} - 1} \to \infty.$$

By Lemma 2, the latter condition is equivalent to

$$\frac{\log n_k}{p_{k-1}} \to \infty$$

and is satisfied if

$$n_k = e^{a_k p_{k-1}},$$

where $\{a_k\}$ is any sequence tending to infinity.

THEOREM 8. *If* $\{n_1, n_2, \ldots\}$ *satisfies* (32), *then there exist functions in* Lip 1 *whose Fourier series with respect to* $\Psi(n_1, n_2, \ldots)$ *are not* $(C,1)$ *summable on a set* S *which is non-denumerable and includes all rationals of the form* r/p_m.

Proof. It will suffice to prove that the assertion is true for the function $x - [x]$. For, as pointed out in the proof of Theorem 5, the expansion of $J_{k-1}(x)$ is essentially the same as that of $x - [x]$.

The latter expansion is given by Theorem 1. Let $s_j(x)$, $\sigma_j(x)$ denote its jth partial sum and $(C, 1)$ mean respectively. Since the series has gaps, $s_j(x)$ will be constant over blocks of consecutive values of j. More precisely,

$$(33)\quad s_j(x) = s_{p_{k-1}}(x) + \frac{1}{p_k}\sum_{r=1}^{\nu}\frac{1}{\omega_k^{-r} - 1}\,\phi_{k-1}^{j}(x),$$

$$\nu p_{k-1} < j \leqslant (\nu+1)p_{k-1};\ 1 \leqslant \nu < n_k.$$

Now suppose $\phi_{k-1}(x) = 1$. It is easy to obtain an expression for $\sigma_j(x)$ from (33) when $j = mp_{k-1}$.

$$(34)\quad mp_{k-1}\,\sigma_{mp_{k-1}}(x) = \sum_{j=1}^{mp_{k-1}} s_j(x) = \sum_{j=1}^{p_{k-1}} s_j(x) + \sum_{j=p_{k-1}+1}^{mp_{k-1}} s_j(x)$$

$$= p_{k-1}\,\sigma_{p_{k-1}}(x) + p_{k-1}(m-1)s_{p_{k-1}}(x) + \frac{p_{k-1}}{p_k}\sum_{r=1}^{m-1}(m-r)\frac{1}{\omega_k^{-r} - 1}.$$

Therefore,

$$(35)\quad \sigma_{mp_{k-1}}(x) = \frac{1}{m}\,\sigma_{p_{k-1}}(x) + \left(1 - \frac{1}{m}\right)s_{p_{k-1}}(x) + \frac{1}{p_k}\sum_{r=1}^{m-1}\left(1 - \frac{r}{m}\right)\frac{1}{\omega_k^{-r} - 1}.$$

We may assume $\{\sigma_{p_{k-1}}(x)\}$ is bounded. Otherwise x is already a point where summability fails. Furthermore, $\{s_{p_{k-1}}(x)\}$ is also bounded. A general theorem of Ohkuma **(3)** says that the p_kth partial sums of the Fourier series of an integrable function converge to the function at all points of continuity. (For $x = 0$, this argument does not apply but a direct verification is easy.) Consequently,

$$(36)\quad \sigma_{mp_{k-1}}(x) = O(1) + \frac{1}{p_k}\sum_{r=1}^{m-1}\left(1 - \frac{r}{m}\right)\frac{1}{\omega_k^{-r} - 1}$$

$$= O(1) + \frac{1}{p_k}\sum_{r=1}^{m-1}\frac{1}{\omega_k^{-r} - 1} - \frac{1}{p_k}\sum_{r=1}^{m-1}\frac{r}{m}\frac{1}{\omega_k^{-r} - 1}.$$

Take $m = [\frac{1}{4}n_k] + 1$. Since $|\omega_k^{-r} - 1|^{-1} = \frac{1}{2}\csc \pi r/n_k < Cn_k/r$ for $0 < r \leqslant [\frac{1}{4}n_k] + 1$, the absolute value of the last sum in (36) is dominated by

$$\frac{1}{p_k}\sum_{r=1}^{m}\frac{r}{m}\frac{Cn_k}{r} = \frac{Cn_k}{p_k} = \frac{C}{p_{k-1}} = o(1).$$

Thus if $\phi_{k-1}(x) = 1$, and $j_k = ([\frac{1}{4}n_k] + 1)p_{k-1}$,

$$(37)\quad \sigma_{j_k}(x) = O(1) + \frac{1}{p_k}\sum_{r=1}^{[n_k/4]}\frac{1}{\omega_k^{-r} - 1}.$$

If for a given x, relation (37) holds for infinitely many values of k, assumption (32) shows that the sequence

$$\{\sigma_{f_k}(x)\}$$

is unbounded. But there is a non-denumerable set S including all rationals r/p_m of such values of x. Let

$$x = \sum_{k=1}^{\infty} \frac{c_k}{p_k}; \quad 0 \leqslant c_k < n_k.$$

Since $\phi_{k-1}(x) = \omega_k{}^{c_k}$, $x \in S$ if and only if infinitely many of the c_k are zero. This completes the proof.

It is known that the Walsh-Fourier series of a function of bounded variation converges at each dyadic rational and at each point of continuity of the function. Since the rationals r/p_m are the analogues of the dyadic rationals, Theorem 8 demonstrates a significant difference between the analytic properties of the Walsh system and those of $\Psi(n_1, n_2, \ldots)$ when $\{n_i\}$ is unbounded.

References

1. H. E. Chrestenson, *A class of generalized Walsh functions*, Pacific J. Math., *5* (1955), 17–31.
2. N. J. Fine, *On the Walsh functions*, Trans. Amer. Math. Soc., *65* (1949), 372–414.
3. T. Ohkuma, *On a certain system of orthogonal step functions*, Tohoku Math. J., *5* (1953), 166–177.
4. N. Vilenkin, *On a class of complete orthonormal systems*, Bull. Acad. Sci. URSS, Ser. Math. (Izvestia Akad. Nauk. SSSR), *11* (1947), 363–400 (Russian, English summary).
5. J. L. Walsh, *A closed set of normal orthogonal functions*, Amer. J. Math., *55* (1923), 5–24.

University of Pennsylvania

THE CLOSED IDEALS IN AN ALGEBRA OF ANALYTIC FUNCTIONS

WALTER RUDIN

I. Introduction. Let K and C be the closure and boundary, respectively, of the open unit disc U in the complex plane. Let $\mathfrak{A}$ be the Banach algebra whose elements are those continuous complex functions on K which are analytic in U, with norm

$$||f|| = \max_{z \in K} |f(z)| \qquad (f \in \mathfrak{A}).$$

Sometimes it will be convenient to say that $f \in \mathfrak{A}$ even if f is defined merely in U but can be extened to K, so that the extended function is a member of $\mathfrak{A}$.

In this paper, all closed ideals of $\mathfrak{A}$ will be determined (Theorem 1). Before the result can be stated, some definitions are required.

A *Blaschke product* is a function of the form

$$B(z) = z^m \prod_{n=1}^{\infty} \frac{a_n - z}{1 - \bar{a}_n z} \cdot \frac{|a_n|}{a_n} \qquad (z \in U), \tag{1.1}$$

where m is a non-negative integer, $0 < |a_n| < 1$, and $\sum\{1 - |a_n|\} < \infty$. The set $\{a_n\}$ may be finite, or even empty.

A *measure* is a complex-valued completely additive set function μ defined for all Borel subsets of C; μ is *singular* if μ is concentrated on a set of Lebesgue measure zero.

Following Beurling **(1**, p. 246**)**, we call a function of the form

$$M(z) = B(z) \exp\left\{ - \int_C \frac{w + z}{w - z} d\lambda(w) \right\} \qquad (z \in U), \tag{1.2}$$

where B is a Blaschke product and λ is a non-negative singular measure, an *inner function*.

We note that a function f, analytic in U, is an inner function if and only if f is bounded in U, f has radial limits of absolute value 1 almost everywhere on C, and the first non-zero Taylor coefficient of f is positive. The necessity of these conditions follows immediately from (1.2). Conversely, if f satisfies these conditions, there is a Blaschke product B such that f/B satisfies the same conditions and has no zeros in U; the function $g = -\log(f/B)$ is analytic in U and has non-negative real part; hence **(5**, p. 185**)** there is a non-negative measure λ such that

$$g(z) = \int_C \frac{w + z}{w - z} d\lambda(w) \qquad (z \in U);$$

Received October 5, 1956. This paper was written while the author was a Research Fellow of the Alfred P. Sloan Foundation.

finally, λ must be singular, since the real part of g has radial limit 0 almost everywhere on C. It follows that f can be written in the form (1.2), i.e., f is an inner function.

If E is a closed subset of C and M is an inner function, we say that M is *associated with* E if all limit points of the zeros of B are in E and if λ is concentrated on E.

If f is bounded and analytic in U, we say that the inner function M *divides* f (or f is *divisible* by M) if f/M is bounded in U. It is shown below (Lemma 7) that if $f \in \mathfrak{A}$ and M divides f, then $f/M \in \mathfrak{A}$.

The main result can now be stated:

THEOREM 1. *Choose a closed subset E of C, of Lebesgue measure zero, and choose an inner function M which is associated with E. Let $I(E, M)$ be the set of all $f \in \mathfrak{A}$ which are divisible by M and which vanish on E. Then $I(E, M)$ is a closed ideal of $\mathfrak{A}$.*

Moreover, every closed ideal of $\mathfrak{A}$ (with the exception of the null ideal) is obtained in this manner.

The restriction on the measure of E is a natural one; for if $f \in \mathfrak{A}$ and $f(z) = 0$ on a set E on C of positive Lebesgue measure, then $f(z) = 0$ for all $z \in K$.

It seems quite remarkable that the inner functions, which are in general discontinuous at the boundary, are found to play such an important role in the ideal structure of $\mathfrak{A}$. Theorem 1 has several consequences which are stated at the end of the paper.

II. Some factorization lemmas. This section contains a number of facts which will be needed in the proof of the main theorem. Although some of these are not new, it seems advisable to include at least sketches of their proofs. In particular, Lemma 2 occurs in **(1**, p. 254**)** as a consequence of an investigation of the Hilbert space H_2. The direct proof given below, based on the purely measure-theoretic Lemma 1, may be of independent interest. Lemma 5 is stated in **(1**, pp. 245–6**)**, but no explicit proof seems to exist in print.

LEMMA 1. *Let Λ be a collection of non-negative measures. There exists a measure λ_0 such that* (a) $0 \leqslant \lambda_0(E) \leqslant \lambda(E)$ *for all $\lambda \in \Lambda$ and all Borel sets E*; (b) *if λ_1 is a measure such that* (a) *is true with λ_1 in place of λ_0, then $\lambda_1(E) \leqslant \lambda_0(E)$ for all Borel sets E.*

We may call λ_0 the largest minorant of Λ.

Proof. Let S be a partition of a fixed Borel set E. That is to say, S is a finite collection of disjoint Borel sets $E_1, \ldots, E_n$ whose union is E. Put

$$\lambda_S(E) = \sum_{i=1}^{n} \inf \lambda(E_i), \qquad (\lambda \in \Lambda)$$

and

$$\lambda_0(E) = \inf \lambda_S(E),$$

the inf being taken over all partitions S of E.

A routine argument shows that the set function λ_0, so defined for all Borel sets, is additive. Complete additivity of λ_0 follows from the fact that λ_0 is majorized by every $\lambda \in \Lambda$. Thus λ_0 is a measure, and it is easy to see that (a) and (b) hold.

LEMMA 2. *Every non-empty collection $\mathfrak{F}$ of inner functions has a greatest common divisor M_0.*

More explicitly, there is an inner function M_0 which divides every $M \in \mathfrak{F}$ (so that M/M_0 is again an inner function), and every inner function which divides every $M \in \mathfrak{F}$ also divides M_0.

Proof. For every $M \in \mathfrak{F}$, write $M = M(B, \lambda)$, where B and λ determine M in accordance with (1.2). Let Λ be the collection of all measures λ which occur in this way, and let λ_0 be the measure whose existence is assured by Lemma 1. Let B_0 be the Blaschke product formed with the zeros which the functions $M \in \mathfrak{F}$ have in common, counting multiplicities. Then $M_0 = M(B_0, \lambda_0)$ is clearly the greatest common divisor of $\mathfrak{F}$, and the Lemma is proved.

We now require another definition. Suppose u is a real function which is summable on C; let β be a real constant; adopting Beurling's terminology **(1, p. 246)** we call the function

$$Q(z) = \exp\left\{\frac{1}{2\pi i}\int_C \frac{w+z}{w-z}\, u(w)\, \frac{dw}{w} + i\beta\right\} \qquad (z \in U) \tag{2.1}$$

an *outer function*. Writing u as the difference of two non-negative summable functions, we see that Q is the quotient of two bounded analytic functions; i.e., Q is of bounded characteristic **(5, p. 178)**.

Since $\log |Q(z)|$ is the Poisson integral of $u(w)$, the radial limits of $\log |Q(z)|$ are equal to $u(w)$ for almost all $w \in C$. Thus any outer function has the representation

$$Q(z) = \exp\left\{\frac{1}{2\pi i}\int_C \frac{w+z}{w-z} \log |Q(w)|\, \frac{dw}{w} + i\beta\right\}. \tag{2.2}$$

According to Nevanlinna **(5, p. 190)**, every[1] function f which is of bounded characteristic in U, can be written (uniquely) in the form

$$f(z) = \frac{B_1(z)}{B_2(z)} \exp\left\{\int_C \frac{w+z}{w-z}\, d\mu(w) + i\beta\right\} \qquad (z \in U), \tag{2.3}$$

where B_1, B_2 are Blaschke products, μ is a real-valued measure on C, and β is a real constant.

If we split off the absolutely continuous part of μ and apply the Jordan decomposition to the singular part σ of μ (i.e., $\sigma = \sigma_1 - \sigma_2$, with $\sigma_1 \geqslant 0$, $\sigma_2 \geqslant 0$, and σ_1, σ_2 mutually singular, we arrive at the following result:

[1] We assume here, and throughout this section, that $f(z)$ is not identically zero.

LEMMA 3. *If f is of bounded characteristic in U, then*

$$f(z) = N_f(z)\, Q_f(z) \qquad (z \in U) \tag{2.4}$$

where N_f is the quotient of two inner functions without common factor[2]*, and Q_f is an outer function given by*

$$Q_f(z) = \exp\left\{\frac{1}{2\pi i}\int_C \frac{w+z}{w-z}\log|f(w)|\,\frac{dw}{w} + i\beta\right\} \qquad (z \in U). \tag{2.5}$$

Let us recall that H_1 is the class of all functions h, analytic in U, for which the integral

$$\int_0^{2\pi} |h(re^{i\theta})|d\theta \qquad (0 < r < 1)$$

is a bounded function of r.

LEMMA 4. *If*

$$\int_0^{2\pi} |f(e^{i\phi})|\, d\theta < \infty$$

and Q_f is defined by (2.5)*, then $Q_f \in H_1$.*

Proof. Let $P_r(\theta)$ be the Poisson kernel

$$P_r(\theta) = \frac{1}{2\pi} \cdot \frac{1-r^2}{1-2r\cos\theta + r^2}.$$

Taking absolute values in (2.5), the well-known inequality between the geometric and arithmetic means yields

$$|Q_f(re^{i\theta})| = \exp\left\{\int_0^{2\pi} P_r(\theta-\phi)\log|f(e^{i\phi})|d\phi\right\} \leq \int_0^{2\pi} P_r(\theta-\phi)\,|f(e^{i\phi})|d\phi,$$

so that

$$\int_0^{2\pi} |Q_f(re^{i\theta})|d\theta \leq \int_0^{2\pi} |f(e^{i\phi})|d\phi,$$

and $Q_f \in H_1$.

LEMMA 5. *If $f \in H_1$, then $f = M_f Q_f$, where M_f is an inner function, and Q_f is given by* (2.5).

That is to say, the denominator of N_f in (2.4) reduces to 1. It is easy to deduce that the Lemma is true if $f \in H_p$, for $0 < p < \infty$, but the case $p = 1$ is the one which is needed later.

The representation of f in the form $M_f Q_f$ will be called the *canonical factorization* of f.

Proof. Put $\log^+ = \max(\log, 0)$, $\log^- = -\min(\log, 0)$.

[2] To say that two inner functions have no common factor means that the constant 1 is the only inner function which divides both of them.

Choose $R < 1$, and replace $f(z)$ by $f(Rz)$. Since $\log |f|$ is subharmonic in U, we have, putting $z = re^{i\theta}$,

$$\log |f(Rz)| \leqslant \int_0^{2\pi} P_r(\theta - \phi) \log^+ |f(Re^{i\phi})| d\phi - \int_0^{2\pi} P_r(\theta - \phi) \log^- |f(Re^{i\phi})| d\phi. \tag{2.6}$$

By virtue of the inequality

$$|\log^+ s - \log^+ t| \leqslant |s - t| \qquad (s, t > 0),$$

and the fact that $f(Re^{i\theta})$ tends to $f(e^{i\theta})$ in the norm of L_1, the first integral in (2.6) tends to

$$\int_0^{2\pi} P_r(\theta - \phi) \log^+ |f(e^{i\phi})| d\phi$$

as $R \to 1$. Fatou's lemma, applied to the second integral in (2.6), then leads to

$$\log |f(z)| \leqslant \int_0^{2\pi} P_r(\theta - \phi) \log |f(e^{i\phi})| d\phi \qquad (z = re^{i\theta}). \tag{2.7}$$

But (2.5) shows that $\log |Q_f(z)|$ is equal to the right member of (2.7). Hence f/Q_f is bounded in U, and the radial limits of f/Q_f have absolute value 1 almost everywhere on C (by (2.5)). It follows that f/Q_f is an inner function, if β is properly chosen in (2.5).

LEMMA 6. *Suppose $f \in \mathfrak{A}$. Let E_f be the set of all $z \in C$ such that $f(z) = 0$. Let $f = M_f Q_f$ be the canonical factorization of f. Then M_f is associated with E_f, $Q_f \in \mathfrak{A}$, and $Q_f(z) = 0$ on E_f.*

Proof. The zeros of M_f in U are the same as those of f, hence their limit points lie in E_f. If the measure λ of (1.2) were not concentrated on E_f, then λ, considered as a function of bounded variation, would have $+\infty$ as derivative at some point $z_0 \in C - E_f$ (7, p. 128); taking absolute values in (1.2), the resulting Poisson integral shows that $M(rz_0) \to 0$ as $r \to 1$. But $|Q_f(z)| \leqslant |f(z)|$, so that Q_f is bounded in U. This implies that $f(z_0) = 0$, a contradiction. Thus M_f is associated with E_f.

It follows that M_f is continuous and different from 0 at every point of $C - E_f$, so that $Q_f = f/M_f$ is continuous on $K - E_f$. The Poisson integral representation of $\log |Q_f(z)|$ (the right member of (2.7)) shows that $Q_f(z) \to 0$ as $z \to z_0 \in E_f$, and the lemma follows.

LEMMA 7. *If $f \in \mathfrak{A}$ and M is an inner function which divides f, then $f/M \in \mathfrak{A}$.*

Proof. Put $g = f/M$, $f = M_f Q_f$, $g = M_g Q_g$. By the uniqueness of the canonical factorizations, $Q_g = Q_f$ and $M_g = M_f/M$. Since M_f is associated with E_f, so is M_g. By Lemma 6, $Q_g \in \mathfrak{A}$ and $Q_g(z) = 0$ on E_f. Hence $M_g Q_g \in \mathfrak{A}$.

III. Proof of Theorem 1. We now assume that E is a closed subset of C, of Lebesgue measure zero, that M is an inner function associated with E, and that $I(E, M)$ is the set of all $f \in \mathfrak{A}$ which are divisible by M and which vanish on E.

It is clear that $I(E,M)$ is an ideal; we have to prove that $I(E,M)$ is closed.

Choose any $g \in \mathfrak{A}$ which is in the closure of $I(E, M)$. Fix r, $0 < r < 1$, such that $M(z) \neq 0$ if $|z| = r$. Put

$$\delta = \min_{\theta} |M(re^{i\theta})|. \tag{3.1}$$

Then $0 < \delta \leqslant 1$. By our choice of g, the definition of $I(E, M)$, and Lemma 7, there exists an $f \in \mathfrak{A}$ such that

$$\| Mf - g \| < \delta. \tag{3.2}$$

This implies

$$\| f \| = \| Mf \| < \| g \| + \delta. \tag{3.3}$$

If $|z| = r$, (3.1) and (3.2) lead to

$$\left| f(z) - \frac{g(z)}{M(z)} \right| < 1, \tag{3.4}$$

so that

$$\left| \frac{g(z)}{M(z)} \right| < 1 + |f(z)| < 1 + \|g\| + \delta \leqslant 2 + \|g\| \qquad (|z| = r). \tag{3.5}$$

Since (3.5) is true for all r such that $M(z) \neq 0$ on $|z| = r$, g/M is bounded in U, so that M divides g. Also, being in the closure of $I(E, M)$, $g(z) = 0$ if $z \in E$. Hence $g \in I(E, M)$, so that $I(E, M)$ is closed, and the first part of Theorem 1 is proved.

To prove the second part, let J be a closed ideal of $\mathfrak{A}$, distinct from the null ideal. Consider the canonical factorizations $f = M_f Q_f$ for all $f \in J$ which do not vanish identically, and let M be the greatest common divisor of the functions M_f so obtained (Lemma 2). Let E_f be the subset of C on which $f(z) = 0$, and let E be the intersection of these sets E_f.

It is then clear that J is contained in $I(E, M)$. We have to prove that these two ideals are actually equal.

Let J_1 be the set of all functions f/M, with $f \in J$. Since M divides M_f, Lemma 7 shows that J_1 is a subset of $\mathfrak{A}$. Hence it is clear that J_1 is an ideal of $\mathfrak{A}$; and since $\| f/M \| = \| f \|$, J_1 is a closed ideal. If we can prove that $J_1 = I(E, 1)$, it will therefore follow that $J = I(E, M)$.

Thus it suffices to prove the theorem for the case in which M is the constant 1, i.e., there is no non-constant inner function which divides every $f \in J$.

Let μ be a measure which annihilates J, in the sense that

$$\int_C f(w) d\mu(w) = 0 \qquad (f \in J) \tag{3.6}$$

Since J is an ideal,

$$\int_C k(w)\, f(w) d\mu(w) = 0 \quad (k \in \mathfrak{A}, f \in J). \tag{3.7}$$

By a well-known theorem of F. and M. Riesz **(6)**, (3.7) implies that for every $f \in J$ there is a function $h_f \in H_1$ whose boundary values give rise to the following equality between measures on C:

$$f(w)\, d\mu(w) = h_f(w)\, dw. \tag{3.8}$$

If S is a closed subset of $C - E_f$, of Lebesgue measure zero, (3.8) shows that the total variation of μ on S is zero. This is true for all $f \in J$, so that the singular part σ of μ is concentrated on E. Put $\alpha = \mu - \sigma$. Then there exists $\phi \in L_1$ on C such that

$$d\alpha(w) = \phi(w)\, dw. \tag{3.9}$$

Since $f(w) = 0$ on E and σ is concentrated on E, we have, by (3.8) and (3.9),

$$h_f(w)\, dw = f(w)\, d\mu(w) = f(w)\, d\alpha(w) = f(w)\, \phi(w)\, dw,$$

so that

$$h_f(w) = f(w)\, \phi(w) \tag{3.10}$$

for all $f \in J$ and almost all $w \in C$. Put

$$g_f(z) = \frac{h_f(z)}{f(z)} \qquad (f \in J, z \in U). \tag{3.11}$$

The functions g_f, meromorphic in U, have the same non-tangential boundary values, namely $\phi(w)$, for almost all $w \in C$. Hence **(4**; p. 159**)** $g_f = g$, a function independent of f. Since the functions $f \in J$ have no common zeros in U, g is analytic in U. Being of bounded characteristic, g is of the form (2.4), i.e., $g = N_g Q_g$, with $Q_g \in H_1$ (Lemma 4), unless g is identically zero.

Writing $h_f = h = M_h Q_h$, $f = M_f Q_f$, (3.11) gives

$$g = N_g Q_g = \frac{M_h Q_h}{M_f Q_f}.$$

It follows that $M_h = M_f N_g$, so that the denominator of N_g divides M_f, for *every* $f \in J$. Since we are dealing with the case $M = 1$, this implies that the denominator of N_g also reduces to 1, so that N_g is an inner function. Hence $g \in H_1$. This is of course true *a fortiori* in the excluded case (g identically zero).

Since g has boundary values $\phi(w)$, we have

$$\int_C k(w)\, \phi(w) dw = 0 \qquad (k \in \mathfrak{A}).$$

Now if $k \in \mathfrak{A}$ and k vanishes on E, that is to say, if $k \in I(E, 1)$, then

$$\int_C k(w) d\mu(w) = \int_C k(w) d\alpha(w) = \int_C k(w)\, \phi(w) dw = 0.$$

Thus every measure which annihilates J also annihilates $I(E, 1)$. Hence J cannot be a proper subspace of $I(E, 1)$, and the theorem is proved.

IV. Consequences of Theorem 1. A closed ideal J in a Banach algebra is said to be *principal* if it is generated by a single element g, i.e., if J is the smallest closed ideal which contains g.

THEOREM 2. *Every closed ideal of $\mathfrak{A}$ is principal.*

Proof. Consider a closed ideal $I(E, M)$. Construct a negative summable function u on C such that u has a bounded derivative on every closed subarc of $C - E$, and such that $u(w) \to -\infty$ as $w \to w_0$, for all $w_0 \in E$. Let Q be the outer function obtained from this u by formula (2.1), with $\beta = 0$. Then $Q \in \mathfrak{A}$ (compare **(2**, pp. 342–4, 360–1**)**).

If $g = MQ$, then $g \in \mathfrak{A}$, since M is associated with E and Q vanishes on E. By Theorem 1, $I(E, M)$ is equal to the closed ideal generated by g, so that $I(E, M)$ is principal. The theorem follows.

Among other things, Theorem 1 implies the well-known theorem that the maximal ideals of $\mathfrak{A}$ consist of all members of $\mathfrak{A}$ which vanish at a given point of K. In fact, $I(E, M)$ is maximal in exactly the following three cases:

(a) E is empty, $M(z) = z$.

(b) E is empty, $M(z) = \dfrac{a - z}{1 - \bar{a}z} \cdot \dfrac{|a|}{a}$ for some $a \in U$, $a \neq 0$.

(c) E consists of one point of C, $M(z) = 1$.

The question often arises in the study of Banach alegbras whether every closed ideal is the intersection of maximal ideals (or whether every closed ideal is the kernel of its hull, in Segal's terminology **(3**, p. 56**)**). Theorem 1 shows that in $\mathfrak{A}$ the following is the case:[3]

THEOREM 3. *$I(E, M)$ is the intersection of maximal ideals of $\mathfrak{A}$ if and only if M is a Blaschke product without multiple zeros; in particular, the measure λ in the representation* (1.2) *of M must be zero.*

A closed ideal in a Banach algebra is said to be primary if it is contained in only one maximal ideal. There are two kinds of primary ideals in $\mathfrak{A}$. First, if $a \in U$ and m is a positive integer, the ideal $J(a, m)$ generated by $(z - a)^m$ is primary; with each $a \in U$ there is thus associated a countable set of primary ideals. Secondly, if $a \in C$ and t is a non-negative real number, the ideal $J(a, t)$ generated by

$$(z - a) \exp\left\{-t \frac{a + z}{a - z}\right\}$$

[3]Whenever the symbol $I(E, M)$ appears, it is understood that M is associated with E.

is primary. For $a \in U$, $J(a, m) \subset J(a, n)$ if and only if $m \geqslant n$; for $a \in C$, $J(a, t) \subset J(a, s)$ if and only if $t \geqslant s$. There are no other inclusion relations between primary ideals.

In some Banach algebras every closed ideal is the intersection of primary ideals, although there are closed ideals which are not intersections of maximal ideals (for references, see (3, p. 182)). Theorem 1 shows that in $\mathfrak{A}$ the situation is as follows:

THEOREM 4. *$I(E, M)$ is the intersection of primary ideals of $\mathfrak{A}$ if and only if the measure λ in the representation (1.2) of M has no continuous component (i.e., if λ is the sum of an at most countable number of point-measures concentrated on E).*

REFERENCES

1. Anne Beurling, *On two problems concerning linear transformations in Hilbert space*, Acta Math., *81* (1949), 239–255.
2. P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Math., *30* (1906), 335–400.
3. L. H. Loomis, *An Introduction to Abstract Harmonic Analysis* (New York, 1953).
4. N. Lusin and J. Priwaloff, *Sur l'unicité et la multiplicité des fonctions analytiques*, Ann. Sci. Ec. Norm. Sup., *42* (1925), 143–191.
5. Rolf Nevanlinna, *Eindeutige analytische Funktionen* (Berlin, 1936).
6. F. and M. Riesz, *Ueber die Randwerte einer analytischen Funktion*, Quatrième congrès des math. scand., 1916, 27–44.
7. Stanislaw Saks, *Theory of the Integral* (Warsaw, 1937).

University of Rochester

REPRESENTATION OF BANACH ALGEBRAS WITH AN INVOLUTION

J. A. SCHATZ

1. Introduction. In 1943 Gelfand and Neumark **(3)** characterized uniformly closed self-adjoint algebras of bounded operators on a Hilbert space as Banach algebras with an involution (a conjugate linear anti-isomorphism of period two) satisfying several additional conditions. The main purpose of this paper is to point out that if we consider algebras of bounded operators on complex Banach spaces more general than Hilbert space, then we can represent a larger class of algebras by essentially the same methods. The Banach spaces considered are those which have an inner product satisfying the algebraic conditions imposed on a Hilbert space inner product except that positivity is replaced by non-degeneracy and continuity of the inner product. We shall call such Banach spaces BIP-spaces. Now let $\mathfrak{A}$ be a complex Banach algebra with an involution satisfying $||a^*|| = ||a||$. Our main result (Theorem 3.6) is that there is a BIP-space $\mathfrak{X}$ and an involution-preserving isomorphism of $\mathfrak{A}$ onto an algebra of bounded operators on $\mathfrak{X}$ which is norm-preserving on self-adjoint elements.

I should like to take this opportunity to thank C. E. Rickart for many helpful suggestions and to express my deep appreciation for his many acts of kindness.

2. BIP-spaces. In this section we shall define BIP-spaces and present those facts which we shall need to state and prove our representation theorem. It is our intention to present a detailed study of BIP-spaces at a later date. Complex conjugates will be indicated by a superscript c.

Definition 2.1. A BIP-space $\mathfrak{X}$ is a complex Banach space which has an inner product (x, y) defined from $\mathfrak{X} \times \mathfrak{X}$ to the complex numbers such that for arbitrary $x, y, z \in \mathfrak{X}$ and complex λ we have

(1) $(x + \lambda y, z) = (x,z) + \lambda(y,z)$,
(2) $(x,y) = (y,x)^c$,
(3) $(x,z) = 0$ for all $z \in \mathfrak{X}$ if and only if $x = 0$,
(4) $|(x,y)| \leqslant ||x||\,||y||$.

As a result of (4), $x_n \to x$ and $y_n \to y$ imply $(x_n, y_n) \to (x, y)$. This joint continuity of the inner product is equivalent to (4) in the sense that in the presence of joint continuity of the inner product $\mathfrak{X}$ may be given an equivalent norm in which (4) holds.

Received April 12, 1956. Portions of a doctoral dissertation submitted to Brown University are contained in this paper.

For a given inner product all norms under which a linear space L becomes a BIP-space are equivalent. This follows from results of Rickart (7, Lemmas 2.1 and 2.4) by noting that for each $x \in L$, $(\cdot, x)$ is a linear functional on L bounded with respect to each such norm and that 0 is the only element at which all of these functionals vanish.

Let S be a set, let $\mathfrak{X}_s$ be a BIP-space for each $s \in S$, and let $\mathfrak{X} = \sum_{s \in S} \mathfrak{X}_s$ be the set of functions x defined on S satisfying $x(s) \in \mathfrak{X}_s$ for each $s \in S$ and $\sum_{s \in S} \|x(s)\|$ exists. Define operations on $\mathfrak{X}$ by $(x + y)(s) = x(s) + y(s)$, $(\lambda x)(s) = \lambda x(s)$, $\|x\| = \sum_{s \in S} \|x(s)\|$ and $(x, y) = \sum_{s \in S} (x(s), y(s))$ for each $x, y \in \mathfrak{X}$ and $s \in S$. Then $\mathfrak{X}$ is a BIP-space which we shall call the direct sum of the $\mathfrak{X}_s$'s.

Let $\mathfrak{X}$ be a BIP-space. By an operator on $\mathfrak{X}$ we shall mean an additive homogeneous function from $\mathfrak{X}$ to $\mathfrak{X}$. If T is an operator which is bounded with respect to the norm on $\mathfrak{X}$ we shall denote the operator bound by $\|T\|$. We shall denote the algebra of all bounded operators on $\mathfrak{X}$ by $B(\mathfrak{X})$. Let T and S be operators on $\mathfrak{X}$. If there is an operator T^* on $\mathfrak{X}$ such that $(Tx, y) = (x, T^*y)$ for all $x, y \in \mathfrak{X}$, then we shall say T^* is adjoint to T. If T has an adjoint, it is unique and we are justified in using the notation T^* to denote it. If T^* exists, then T is automatically bounded.[1] This follows from the closed graph theorem (4, Theorem 2.13.9) by noting that if $x_n \to x$ and $\lim Tx_n$ exists, then $Tx = \lim Tx_n$ since

$$(Tx - \lim Tx_n, y) = (x, T^*y) - \lim (x_n, T^*y) = 0.$$

The following facts are easy to check. If T^* and S^* exist then for complex λ, $(T + \lambda S)^*$ exists and is equal to $T^* + \lambda^c S^*$. If T^* exists then $(T^*)^*$ exists and is equal to T. If T^* and S^* exist, then $(TS)^*$ exists and is equal to S^*T^*. The identity operator, which we shall always denote by I, has an adjoint and $I^* = I$. An algebra of operators on $\mathfrak{X}$ is called self-adjoint if and only if for each T in the algebra, T^* exists and is in the algebra.

The following lemma is a modification of a result of Lax (5).

LEMMA 2.2. *Let $\mathfrak{X}$ be a* BIP*-space. If $(x,x) > 0$ for all nonzero $x \in \mathfrak{X}$ and T is an operator on $\mathfrak{X}$ such that T^* exists, then*

$$\sup_{x \neq 0} \frac{(Tx, Tx)}{(x, x)} \leqslant \|T^*T\|.$$

Proof. The usual calculation for Hilbert space shows that if $y, z \in \mathfrak{X}$, then $|(y,z)|^2 \leqslant (y,y)(z,z)$. Therefore for any operator S which has an adjoint,

$$(Sx, Sx)^2 = (S^*Sx, x)^2 \leqslant (S^*Sx, S^*Sx)(x,x).$$

Noting that $(S^*S)^* = S^*S$ and applying this repeatedly we see that for any positive integer n,

[1] This result was noted by the referee. It allows us to remove an unnecessary boundedness condition which was originally present in Lemma 2.2.

$$(Tx, Tx)^{2^n} \leqslant ((T^*T)^{2^{n-1}} x, (T^*T)^{2^{n-1}} x)\,(x, x)^{2^n-1}$$
$$= ((T^*T)^{2^n} x, x)\,(x, x)^{2^n-1} \leqslant ||T^*T||^{2^n}\,||x||^2\,(x, x)^{2^n-1}.$$

Therefore,

$$\frac{(Tx, Tx)}{(x, x)} \leqslant ||T^*T||\left(\frac{||x||^2}{(x, x)}\right)^{2^{-n}}.$$

The proof is completed by noting that

$$\lim_{n\to\infty}\left(\frac{||x||^2}{(x, x)}\right)^{2^{-n}} = 1.$$

3. *-algebras and the representation theorem.

Definition 3.1. By a **-algebra* $\mathfrak{A}$ we shall mean a complex Banach algebra with identity e, $||e|| = 1$, and a function $a \to a^*$ from $\mathfrak{A}$ into $\mathfrak{A}$ such that for all $a,b \in \mathfrak{A}$ and complex λ we have,

(i) $(a + \lambda b)^* = a^* + \lambda^* b^*$,
(ii) $(ab)^* = b^*a^*$,
(iii) $(a^*)^* = a$,
(iv) $||a^*|| = ||a||$.

A function satisfying (i) to (iii) is called an *involution*. It is apparent that continuity of the involution follows from (iv). Conversely if a complex Banach algebra with e, $||e|| = 1$, has a continuous involution and we replace the norm by the equivalent norm $||a||' = \max(||a||, ||a^*||)$ then we obtain a *-algebra.

The extensiveness of *-algebras may be judged from the fact that any Banach algebra $\mathfrak{B}$ can be embedded in a *-algebra $\mathfrak{B}'$. To see this, let $\mathfrak{B}'$ be $\mathfrak{B} \times \mathfrak{B}$ with

$$(x,y) + (x',y') = (x + x', y + y'),$$
$$\lambda(x,y) = (\lambda x, \lambda^* y),\ (x,y)(x'y') = (xx', y'y),$$
$$||(x,y)|| = \max(||x||, ||y||),\ (x,y)^* = (y,x).$$

An element $h \in \mathfrak{A}$ for which $h^* = h$ is called self-adjoint. It is easy to see that e and 0 are self-adjoint and so are all elements of the form a^*a, aa^*, and $a + a^*$. Let H be the set of all self-adjoint elements of $\mathfrak{A}$. Then H is a real Banach space. For $a \in \mathfrak{A}$ there are unique self-adjoint elements h,k such that $a = h + ik$; in fact, $h = \frac{1}{2}(a + a^*)$ and $k = -\frac{1}{2}i(a - a^*)$. If $a = h + ik$ with $h, k \in H$ we shall call h and k the self-adjoint components of a.

We shall call a linear functional f on $\mathfrak{A}$ (1) a real functional if and only if $f(a^*) = f(a)^*$ for all $a \in \mathfrak{A}$, and (2) a positive functional if and only if $f(a^*a) \geqslant 0$ for all $a \in \mathfrak{A}$. A positive functional is necessarily real and bounded (by $f(e)$) **(6**, pp. 29–32**)**. The extensiveness of real functionals is indicated by the following facts. If f' is a (real) linear functional on H, then there is a unique linear functional f on $\mathfrak{A}$ such that f restricted to H is equal to f'; in fact,

$f(a) = f'(h) + if'(k)$ for $a = h + ik$ with $h, k \in H$. Moreover f is bounded if and only if f' is. A linear functional on $\mathfrak{A}$ is real if and only if it assumes real values on H. If f is a linear functional on A then there are unique real functionals f_1, f_2 such that $f = f_1 + if_2$; in fact,

$$f_1(a) = \tfrac{1}{2}[f(a) + f(a^*)^-], \; f_2(a) = -\tfrac{1}{2}i[f(a) - f(a^*)^-].$$

Moreover f is bounded if and only if both f_1 and f_2 are. A *-algebra need have no (nonzero) positive functionals. To see this consider $\mathfrak{B}'$ as above. Since $(e, -e)^*(e, -e) = -(e, e)$ and (e, e) is the identity for $\mathfrak{B}'$, every positive functional must vanish at (e, e) and therefore is identically zero.

Let f be a real functional on $\mathfrak{A}$ and let $N(f)$ be the set of all $a \in \mathfrak{A}$ such that $f(ba) = 0$ for all $b \in \mathfrak{A}$. Then $N(f)$ is a left ideal. If f is bounded then $N(f)$ is closed and the difference space $\mathfrak{A} - N(f)$ (with elements a_f for $a \in \mathfrak{A}$) is a Banach space under the norm

$$\inf_{b \in N(f)} \|a + b\|.$$

We shall call $N(f)$ the reducing ideal of f.

Definition 3.2. Suppose that $a \to T_a$ is a homomorphism of A into an algebra of operators on a BIP-space. Then we shall call the homomorphism a *-homomorphism if and only if for all $a \in \mathfrak{A}$, $T_{a^*} = T_a^*$ and $T_e = I$.

LEMMA 3.3. *Let f be a real functional on $\mathfrak{A}$ with $\|f\| \leqslant 1$ and let $N(f)$ be the reducing ideal of f. Then $\mathfrak{A} - N(f)$ is a* BIP*-space and there is a norm-reducing *-homomorphism of $\mathfrak{A}$ onto a self-adjoint algebra of bounded operators on $\mathfrak{A} - N(f)$. The kernel of this homomorphism is contained in $N(f)$.*

Proof. For $a, b \in \mathfrak{A}$, we define (a_f, b_f) to be $f(b^*a)$. If $c, d \in N(f)$, then

$$f((b + d)^*(a + c)) = f(b^*a) + f(a^*d)^- + f((b + d)^*c) = f(b^*a).$$

Therefore (a_f, b_f) is independent of the representatives from a_f and b_f.

Since f is linear,

$$(a_f + \lambda b_f, c_f) = f(c^*a) + \lambda f(c^*b) = (a_f, c_f) + \lambda(b_f, c_f).$$

Since f is a real functional,

$$(a_f, b_f) = f(a^*b)^- = (b_f, a_f)^-.$$

If $(a_f, b_f) = 0$ for all $b \in \mathfrak{A}$, then $f(ca) = 0$ for all $c \in \mathfrak{A}$ and consequently $a_f = 0$. If $c \in a_f$ and $d \in b_f$, then

$$|(a_f, b_f)| = |f(d^*c)| \leqslant \|d\| \, \|c\|.$$

Therefore

$$|(a_f, b_f)| \leqslant \|a_f\| \, \|b_f\|.$$

Since $\mathfrak{A} - N(f)$ is a Banach space, it follows that it is a BIP-space.

For $b_f = c_f$, we have $(ab)_f = (ac)_f$ since $N(f)$ is a left ideal. Therefore we can define an operator on $\mathfrak{A} - N(f)$ by $T_a(b_f) = (ab)_f$ for each $a \in \mathfrak{A}$.

Since

$$||(ab)_f|| = \inf_{c\in N(f)}||ab + c|| \leqslant \inf_{c\in N(f)}||a(b + c)|| \leqslant ||a||\inf_{c\in N(f)}||b + c|| = ||a||\,||b_f||,$$

we conclude that $||T_a|| \leqslant ||a||$. Since

$$(T_a(b_f), c_f) = f(c^*ab) = f((a^*c)^*b) = (b_f, T_{a^*}(c_f))$$

we conclude that the mapping $a \to T_a$ is a *-homomorphism.

Finally if a is not in $N(f)$, then $T_a(e_f) = a_f \neq 0$ and therefore $T_a \neq 0$. This means that the kernel of the homomorphism is contained in $N(f)$ which concludes the proof.

When necessary to avoid confusion we shall denote T_a by T_{af}.

LEMMA 3.4. *Let S be a set of real functionals f satisfying $||f|| \leqslant 1$. Let*

$$\mathfrak{X} = \sum_{f\in S} \mathfrak{A} - N(f).$$

*Then there is a norm-reducing *-homomorphism of A onto a self-adjoint algebra of bounded operators on $\mathfrak{X}$ with kernel contained in $\bigcap_{f\in S}N(f)$.*

Proof. For $a \in \mathfrak{A}$, define T_a on $\mathfrak{X}$ by $(T_a\, x)\,(f) = T_{af}(x(f))$ for all $f \in S$, $x \in \mathfrak{X}$. A straightforward check shows that $a \to T_a$ satisfies the stated conditions.

THEOREM 3.5. *If $\mathfrak{A}$ has an essential involution (i.e., for each $a \in \mathfrak{A}$ there is a positive functional f with $f(e) \leqslant 1$ and $f(a^*a) > 0$), then there is a norm-reducing *-isomorphism of A onto a self-adjoint algebra of bounded operators on a Hilbert space.*

Proof. Let S in Lemma 3.4 be the set of all positive functionals f satisfying $f(e) \leqslant 1$. Then the mapping $a \to T_a$ of Lemma 3.4 is a norm-reducing *-isomorphism of A onto a self-adjoint algebra of bounded operators on a BIP-space $\mathfrak{X}$. Moreover, the inner product in $\mathfrak{X}$ is positive definite. By Lemma 2.2,

$$\sup_{x\neq 0}\,(T_a\, x,\, T_a\, x)/(x,x) \leqslant ||T_a^*T_a|| \leqslant ||a^*a|| \leqslant ||a||^2.$$

Therefore each T_a can be extended to a bounded operator on the completion of $\mathfrak{X}$ in the Hilbert space norm.

THEOREM 3.6. *Let $\mathfrak{A}$ be a *-algebra. Then there is a* BIP*-space $\mathfrak{X}$ and a norm-reducing *-isomorphism $a \to T_a$ of $\mathfrak{A}$ onto a self-adjoint algebra of bounded operators on $\mathfrak{X}$ which satisfies $||T_h|| = ||h||$ for each self-adjoint $h \in \mathfrak{A}$.*

Proof. Suppose that h is self-adjoint. Let g be a linear functional on $\mathfrak{A}$ satisfying $||g|| = 1$, $g(h) = ||h||$. Let $f(a) = \frac{1}{2}(g(a) + g(a^*)^c)$ for each $a \in \mathfrak{A}$. Then f is a real functional and $||f|| \leqslant 1$ since

$$|f(a)| \leqslant \tfrac{1}{2}(||a|| + ||a^*||) = ||a||.$$

For arbitrary $a \in N(f)$, we have $||h|| = f(h) = f(h + a) \leqslant ||h + a||$. Therefore $||h_f|| = ||h||$. Let S in Lemma 3.4 be the set of all real functionals with norm at most 1. Define $x \in \mathfrak{X}$ by

$$x(f) = e_f, \quad x(f') = 0, \qquad f' \in S, f' \neq f.$$

Then $||x|| \leqslant 1$ and $||T_h|| \geqslant ||T_h(x)|| = ||T_{h,f}(e_f)|| = ||h_f|| = ||h||$. Therefore $a \to T_a$ is norm-preserving on self-adjoint elements and all that remains is to prove that it is 1:1.

To this end suppose $T_a = 0$. Let $a = h + ik$ with h, k self-adjoint. Then $T_{2h} = T_a + T_a^* = 0$ and $T_{2k} = T_{-ia} + T_{-ia}^* = 0$. Therefore $h = k = 0$ and consequently $a = 0$.

Remark 1. The restriction that $\mathfrak{A}$ has an identity element can be removed. This can be seen by using the standard imbedding of $\mathfrak{A}$ in an algebra $\mathfrak{B}$ with an identity element. $\mathfrak{B}$ consists of all ordered pairs (λ, a) for λ complex and $a \in \mathfrak{A}$. The operations in $\mathfrak{B}$ are defined as if (λ, a) were $\lambda + a$, $(\lambda, a)^* = (\lambda^*, a^*)$, and $||(\lambda, a)|| = |\lambda| + ||a||$.

Remark 2. We can remove the requirement that $\mathfrak{A}$ be complete if we relinquish the requirement that $\mathfrak{X}$ be complete.

Remark 3. A purely algebraic analogue of Theorem 3.6 holds. In this case we let $\mathfrak{X}$ be the finite direct sum of the $\mathfrak{A} - N(f)$'s for all real functionals f.

Remark 4. Finally we indicate a method of obtaining the Gelfand-Neumark result from our considerations. Let $\mathfrak{A}$ be a C^*-algebra (i.e., $\mathfrak{A}$ is a *-algebra, $||a^*a|| = ||a||^2$, and $(e + a^*a)^{-1}$ exists for each $a \in \mathfrak{A}$). It is known that this last condition is superfluous **(8)**. The proof consists of (1) showing that the involution in $\mathfrak{A}$ is essential, (2) appealing to Theorem 3.5, and (3) showing that the isomorphism of Theorem 3.5 is automatically norm-preserving. Except for a minor modification in the method of showing that the T_a's are bounded this is the Gelfand-Neumark proof.

The necessity of demonstrating (3) can be removed if we strengthen (1). Suppose $a \in \mathfrak{A}$. Fukamiya **(2**, Lemma 5**)** has shown that there is a positive functional f satisfying $f(e) = 1$, $f(a^*a) = ||a||^2$. Define $x \in \mathfrak{X}$ by $x(f) = e_f$ and $x(f') = 0$ for $f' \in S, f' \neq f$. Then

$$\sup_{y \neq 0} ((T_a y, T_a y)/(y,y)) \geqslant (T_a x, T_a x)$$
$$= (T_{a_f} e_f, T_{a_f} e_f) = (a_f, a_f) = f(a^*a) = ||a||^2.$$

Therefore the mapping of Theorem 3.5 is norm-preserving.

4. Representations. A preliminary study of *-algebras indicates that many representations are pathological. We shall limit ourselves to the consideration of a more tractable case. Throughout this section $\mathfrak{A}$ will be a *-algebra satisfying $||a^*a|| \geqslant K ||a||^2$ for some $K > 0$. By a representation of $\mathfrak{A}$ we shall

mean a continuous *-homomorphism $a \to T_a$ of $\mathfrak{A}$ onto an algebra of bounded operators on a BIP-space. We shall call the representation irreducible if and only if X has no proper subspaces invariant under each T_a. Finally we shall call a set of representations complete if and only if for each $0 \neq b \in \mathfrak{A}$ there is a representation $a \to T_a$ in the set for which $T_b \neq 0$.

LEMMA 4.1. *If L is a proper left ideal in $\mathfrak{A}$, then the closure of $L + L^*$ does not contain e.*

Proof. By using a result of Arens **(1**, Theorem 1**)** concerning the representation of closed commutative *-subalgebras of $\mathfrak{A}$, it is easy to see that (1) every self-adjoint element has a real spectrum and (2) if $h \in H$ and $\|h - e\| < 1$ then there is a $k \in H$ such that k^{-1} exists, $kh = hk$, and $k^2 = h$.

Now suppose e is in the closure of $L + L^*$. Then there are $x_n, y_n \in L$ such that $x_n + y_n^* \to e$. Then $z_n = \frac{1}{2}(x_n + y_n) \in L$ and $z_n + z_n^* \to e$. Pick m such that $\|e - (z_m + z_m^*)\| < 1$. Let $z_m = h_1 + ih_2$ with $h_1, h_2 \in H$. Let $h = 2h_1 = z_m + z_m^*$. Pick k as in (2). Since $k^{-1}(2h_2)k^{-1} \in H$ we conclude by (1) that $ie - k^{-1}(2h_2)k^{-1}$ has an inverse. It follows from this that

$$(k(e + ik^{-1}(2h_2)k^{-1})k)^{-1}$$

exists. However,

$$k(e + ik^{-1}(2h_2)k^{-1})k = k^2 + i2h_2 = 2(h_1 + ih_2) = 2z_m.$$

Therefore z_m^{-1} exists which is a contradiction since $z_m \in L$.

LEMMA 4.2. *If L is a proper left ideal in $\mathfrak{A}$, then there is a nonzero bounded real functional f on $\mathfrak{A}$ such that $L \subseteq N(f)$.*

Proof. $(L + L^*) \cap H$ is a (real) subspace of H which is non-dense by Lemma 4.1. Using the Hahn-Banach theorem, pick a nonzero bounded (real) linear functional on H which vanishes on $(L + L^*) \cap H$. Let f be the extension of this functional to $\mathfrak{A}$. For $a \in L$, the self-adjoint components of a are in $(L + L^*) \cap H$. Therefore $f(a) = 0$. Since L is a left ideal, we conclude that $L \subseteq N(f)$.

It should be noted that if $\mathfrak{A}$ were a C^*-algebra then we could require f to be a positive functional **(6**, p. 86**)**.

THEOREM 4.3. *$\mathfrak{A}$ has a complete set of irreducible representations.*

Proof. $\mathfrak{A}$ is semi-simple **(7**, Theorem 5.2**)**. Therefore the intersection of the set $\mathscr{M}$ of maximal left ideals in $\mathfrak{A}$ is 0 **(4**, Corollary 1, p. 485**)**. Suppose $0 \neq b \in \mathfrak{A}$. Pick $M \in \mathscr{M}$ such that $b \notin M$. By Lemma 4.2, there is a bounded real function f on $\mathfrak{A}$ such that $M = N(f)$. The representation $a \to T_a$ associated with f is irreducible since a subspace of $\mathfrak{A} - N(f)$ invariant under each T_a must come from a left ideal of $\mathfrak{A}$ containing M. Since the kernel of this representation is contained in M, $T_b \neq 0$.

References

1. R. Arens, *Representation of *-algebras*, Duke Math. J., *14* (1947), 269–282.
2. M. Fukamiya, *On B^*-algebra*, Proc. Japan Acad., *27* (1951), 321–327.
3. I. Gelfand and M. Neumark, *On the imbedding of normed rings into the ring of operators in Hilbert space*, Matematičeskiĭ Sbornik, N.S., *12* (*54*) (1943), 197–217.
4. E. Hille, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium Publications, *31* (New York, 1948).
5. P. D. Lax, *Symmetrizable linear transformations*, Communications on Pure and Applied Mathematics, *7* (1954), 633–647.
6. M. Neumark, *Rings with involution*, Uspehi Matematičeskiĭ Nauk, N.S., *3*, *5* (*27*) (1948), 52–145; translated as Amer. Math. Soc. Translation *25* (1950). All specific references are to the translation.
7. C. E. Rickart, *The uniqueness of norm problem in Banach algebras*, Ann. Math. (2), *51* (1950), 615–628.
8. Review of a paper by Fukamiya in Math. Rev. *14* (1953), 884 and *15* (1954), 1139.

University of Connecticut

ON COUNTABLY PARACOMPACT NORMAL SPACES

M. J. MANSFIELD

1. Introduction. A. H. Stone **(9)**, E. Michael **(3, 4)**, J. L. Kelley and J. S. Griffen **(2)** have established many necessary and sufficient conditions that a regular Hausdorff space be paracompact. It is the purpose of this note to show that if the word "countable" is inserted in the appropriate places in the above-mentioned conditions they become, in general, necessary and sufficient conditions that a normal space be countably paracompact. Making use of one of these conditions, countably paracompact normal spaces are identified as a natural generalization of fully normal spaces.

The following terminology will be used throughout this note. Let X be a topological space. A collection $\mathfrak{R}$ of subsets of X is said to be a *covering* of X if $\bigcup\{R : R \in \mathfrak{R}\} = X$. If $\mathfrak{R}$ and $\mathfrak{S}$ are coverings of X, then $\mathfrak{S}$ is a *refinement* of $\mathfrak{R}$ if each member of $\mathfrak{S}$ is a subset of some member of $\mathfrak{R}$. A collection $\mathfrak{R}$ of subsets of X is *closure-preserving* if for each $\mathfrak{S} \subset \mathfrak{R}$ we have

$$\bigcup\{S^- : S \in \mathfrak{S}\} = (\bigcup\{S : S \in \mathfrak{S}\})^-.$$

The collection $\mathfrak{R}$ is *locally finite* if each $x \in X$ has a neighborhood meeting only a finite number of members of $\mathfrak{R}$; *star-finite* if each member of $\mathfrak{R}$ meets only a finite number of members of $\mathfrak{R}$; *discrete* if each $x \in X$ has a neighborhood meeting at most one member of $\mathfrak{R}$. $\mathfrak{R}$ is *σ-closure-preserving* (resp. *σ-locally finite*, resp. *σ-discrete*) if $\mathfrak{R}$ is the union of countably many closure-preserving (resp. locally finite, resp. discrete) subcollections.

An open covering $\mathfrak{R}$ of a topological space X is said to be *even* if there is a neighborhood U of the diagonal[1] in the cartesian product space $X \times X$ such that for each $x \in X$ the set

$$U[x] = \{y \in X : (x, y) \in U\}$$

is a subset of some member of $\mathfrak{R}$. If $\mathfrak{R}$ and $\mathfrak{S}$ are open coverings of X, then $\mathfrak{S}$ is said to be a *star refinement* of $\mathfrak{R}$ if for each $x \in X$ the set

$$\mathrm{St}(x, \mathfrak{S}) = \bigcup\{S \in \mathfrak{S} : x \in S\}$$

is a subset of some member of $\mathfrak{R}$.[2]

Received August 15, 1956. Presented to the American Mathematical Society, April 14' 1956. This note is based on part of a thesis written by the author under the direction of Professor Melvin Henriksen and accepted by Purdue University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[1]The *diagonal* in $X \times X$ is the set $\{(x, x) : x \in X\}$.

[2]Our usage of "star refinement," which follows **(2**, Exercise 5.U**)**, differs from that of Tukey **(10)**. Tukey refers to our "star refinement" as a "Δ-refinement," reserving the words "star refinement" to describe a slightly different concept.

Finally, a topological space X is said to be *countably paracompact* if each countable open covering of X admits a (countable) locally finite open refinement. A topological space X is *fully normal* if each open covering of X admits an open star refinement.

Throughout the sequel the letter i, j, or k will always stand for a variable which runs over all the natural numbers.

2. The Theorem. We shall frequently need the following two well-known theorems. Theorem 1 is due to Dowker **(1**, Theorem 2**)**; Theorem 2 is due to Morita **(6**, Theorem 3**)**. The reader is referred to the original papers for the proofs.

THEOREM 1 (Dowker). *The following properties of a normal*[3] *space X are equivalent*:

(a) *The space X is countably paracompact.*

(b) *Every countable open covering $\{U_i\}$ of X admits an open refinement $\{V_i\}$ with $\overline{V_i} \subset U_i$.*

(c) *Given a decreasing sequence $\{F_i\}$ of closed sets with a vacuous intersection, there is a sequence $\{G_i\}$ of open sets with vacuous intersection such that $F_i \subset G_i$.*

THEOREM 2 (Morita). *If $\mathfrak{G} = \{G_i\}$ is a countable open covering of a normal space X, and if there exists a closed covering $\mathfrak{F} = \{F_i\}$ of X such that $F_i \subset G_i$, then $\mathfrak{G}$ admits a countable, star-finite, open, star-refinement.*

In his statement of Theorem 2, Morita has the additional hypothesis that X be a T_1-space; this hypothesis, however, is not used in the proof.

Remark. It follows at once from Theorem 2 and condition (b) of Theorem 1 that a normal space X is countably paracompact if and only if each countable open covering of X admits a (countable) star-finite open refinement.

We are now ready to state and prove the main theorem of this note. It should be remarked that some of the implications involved in the theorem are well known.

THEOREM 3. *If X is a normal space, then the following statements are equivalent.*

(a) *The space X is countably paracompact.*

(b) *Each countable open covering of X admits a countable, locally finite, closed refinement.*

(c) *Each countable open covering of X admits a countable, closure-preserving, closed refinement.*

(d) *Each countable open covering of X admits a σ-discrete closed refinement.*

(e) *Each countable open covering of X admits a σ-locally finite closed refinement.*

[3]In our terminology a normal space need not be a Hausdorff space.

(f) *Each countable open covering of X admits a σ-closure-preserving closed refinement.*

(g) *Each countable open covering of X admits a countable, open, star refinement.*

(h) *Each countable open covering of X is even.*

(i) *For each countable locally finite collection $\{A_i\}$ of subsets of X there is a countable, locally finite collection $\{G_i\}$ of open subsets of X such that $A_i \subset G_i$.*

Proof. The pattern of proof is (a) → (b) → (c) → (d) → (e) → (f) → (g) → (h) → (a) → (i) → (a).

(a) *implies* (b): Let $\mathfrak{R} = \{R_i\}$ be a countable open covering of X. We shall show that $\mathfrak{R}$ admits a countable, locally finite, closed refinement. Since X is countably paracompact, there is a locally finite open refinement $\mathfrak{S} = \{S_i\}$ of $\mathfrak{R}$ such that $S_i \subset R_i$ (**1**, proof of Theorem 1). By Theorem 1 (b) there is an open covering $\mathfrak{T} = \{T_i\}$ of X such that $T_i^- \subset S_i$. Evidently $\mathfrak{F} = \{T_i^-\}$ is a countable, locally finite, closed refinement of $\mathfrak{R}$.

(b) *implies* (c) *and* (e) *implies* (f): A locally finite collection of subsets is closure-preserving.

(c) *implies* (d): Any countable collection of subsets is σ-discrete.

(d) *implies* (e): A discrete collection of subsets is locally finite.

(f) *implies* (g): Let $\mathfrak{R} = \{R_i\}$ be a countable open covering of X. We shall show that $\mathfrak{R}$ admits a countable, open, star refinement. By (f), $\mathfrak{R}$ admits a σ-closure-preserving closed refinement $\mathfrak{S} = \bigcup \mathfrak{S}_i$. For each pair of positive integers i, j let

$$T_{i,j} = \bigcup \{S \in \mathfrak{S}_j : S \subset R_i\}.$$

Since each $S \in \mathfrak{S}$ is closed and $\mathfrak{S}_j$ is closure-preserving for each j, we conclude that $T_{i,j}$ is closed.

For each i and any j, $T_{i,j} \subset R_i$. Moreover if $x \in X$, then, since $\mathfrak{S}$ is a covering, there is an integer j and a set $S \in \mathfrak{S}_j$ such that $x \in S$. Since $\mathfrak{S}$ is a refinement of $\mathfrak{R}$, there is an integer i such that $S \subset R_i$. Therefore, $x \in T_{i,j}$. Consequently $\mathfrak{T} = \{T_{i,j}\}$ is a closed refinement of $\mathfrak{R}$.

The family $\mathfrak{T}$ is countable and therefore can be indexed by the natural numbers, say $\mathfrak{T} = \{T_k\}$. Since $\mathfrak{T}$ is a refinement of $\mathfrak{R}$, for each positive integer k there is an integer i_k such that

$$T_k \subset R_{i_k}.$$

Let

$$U_k = R_{i_k}$$

and $\mathfrak{U} = \{U_k\}$. Then $\mathfrak{U}$ is a countable open covering of the normal space X, and $\mathfrak{T}$ is a countable closed covering of X such that $T_k \subset U_k$. Therefore, by Theorem 2, $\mathfrak{U}$ admits a countable, open, star refinement $\mathfrak{V}$. Obviously $\mathfrak{V}$ is also a star refinement of $\mathfrak{R}$.

(g) *implies* (h): We shall show that if an open covering $\mathfrak{R}$ of a topological space X admits an open star refinement, then $\mathfrak{R}$ is an even covering (**2**, Exercise 5.U). Let $\mathfrak{S}$ be an open star refinement of $\mathfrak{R}$, and put $V = \bigcup\{S \times S : S \in \mathfrak{S}\}$. Clearly V is a neighborhood of the diagonal in $X \times X$.

If $x \in X$, choose $R_x \in \mathfrak{R}$ such that $St(x, \mathfrak{S}) \subset R_x$. If $y \in V[x]$ then, by definition of V, there is an $S \in \mathfrak{S}$ such that $x \in S$ and $y \in S$. Therefore $y \in St(x, \mathfrak{S}) \subset R_x$. Hence $V[x] \subset R_x$. Consequently $\mathfrak{R}$ is an even covering of X.

(h) *implies* (a): Let $\mathfrak{R} = \{R_i\}$ be a countable open covering of X. We shall show that $\mathfrak{R}$ admits a star-finite open refinement. It will then follow from the *Remark* above that X is countably paracompact.

By hypothesis, there is a neighborhood U of the diagonal in $X \times X$ such that $\{U[x] : x \in X\}$ is a refinement of $\mathfrak{R}$. Let

$$U^{-1} = \{(x, y) \in X \times X : (y, x) \in U\},$$

and let $V = U \cap U^{-1}$. Evidently V is a *symmetric* neighborhood of the diagonal in $X \times X$; i.e., V is a neighborhood of the diagonal with the property that $(x, y) \in V$ if and only if $(y, x) \in V$.

For each positive integer i let $S_i = \{x \in X : V[x] \subset R_i\}$. Evidently $\{S_i\}$ is a refinement of $\mathfrak{R}$ such that $S_i \subset R_i$. If $x \in S_i^-$, then, since $V[x]$ is a neighborhood of x, $V[x] \cap S_i \neq \phi$, say $y \in V[x] \cap S_i$. Since $y \in S_i$, $V[y] \subset R_i$; but since $y \in V[x]$ and V is symmetric, we conclude that $x \in V[y] \subset R_i$. Hence $S_i^- \subset R_i$. By Theorem 2, $\mathfrak{R}$ admits a star-finite open refinement. Hence X is countably paracompact.

(a) *implies* (i): Let $\{A_i\}$ be a locally finite collection of subsets of X. We must find a locally finite collection $\{G_i\}$ of open sets such that $A_i \subset G_i$.

Let $F_k = \bigcup\{A_p^- : p \geqslant k\}$. Since $\{A_i\}$ is locally finite, F_k is closed for each k. Moreover $F_k \supset F_{k+1}$ and $A_k^- \subset F_k$.

Now if $x \in X$, then x belongs to only a finite number of the A_i^-, say

$$A_{i_1}^-, \ldots, A_{i_n}^-.$$

Set

$$m_x = 1 + \max\{i_1, \ldots, i_n\}.$$

Then for $k \geqslant m_x$, $x \notin A_k^-$. Hence

$$x \notin F_{m_x}.$$

Therefore $\bigcap F_k = \phi$.

Since X is countably paracompact and normal, we may apply Theorem 1 (c): there is a sequence $\{H_k\}$ of open sets such that $F_k \subset H_k$ for each k and $\bigcap H_k = \phi$.

Now $F_1 \subset H_1$. Since X is normal, there is an open set G_1 such that $F_1 \subset G_1 \subset G_1^- \subset H_1$. Proceeding inductively, suppose that for each positive integer $p \leqslant n$ open sets G_p have been defined such that $G_1 \supset G_2 \supset \ldots \supset G_n$, and such that

$$F_p \subset G_p \subset G_p^- \subset H_p \qquad (p = 1, 2, \ldots, n).$$

Now $F_{n+1} \subset F_n \subset G_n$ and $F_{n+1} \subset H_{n+1}$. Therefore $F_{n+1} \subset G_n \cap H_{n+1}$. By the normality of X, there is an open set G_{n+1} such that

$$F_{n+1} \subset G_{n+1} \subset \bar{G}_{n+1} \subset G_n \cap H_{n+1}.$$

Consequently $F_{n+1} \subset G_{n+1} \subset \bar{G}_{n+1} \subset H_{n+1}$ and $G_{n+1} \subset G_n$.

We shall now show that $\{G_k\}$ is locally finite. If $x \in X$, then, since $\bigcap H_k = \phi$, there is an n_x such that

$$x \notin H_{n_x}.$$

Hence, since

$$H_{n_x} \supset \bar{G}_{n_x}, \quad x \notin \bar{G}_{n_x}.$$

Therefore there is a neighborhood V_x of x such that

$$V_x \cap G_{n_x} = \phi.$$

Consequently, if $k > n_x$, then $V_x \cap G_k = \phi$, because

$$G_k \subset G_{n_x}.$$

Therefore V_x meets at most

$$G_1, \ldots, G_{n_x - 1}.$$

Hence $\{G_k\}$ is locally finite.

Finally, for each k, $A_k \subset \bar{A}_k \subset F_k \subset G_k$.

(i) *implies* (a): Let $\mathfrak{R} = \{R_i\}$ be a countable open covering of an arbitrary space X. Put $S_i = \bigcup\{R_p : p = 1, \ldots, i\}$; let $A_1 = S_1$ and $A_i = S_i - S_{i-1}$ $(i = 2, 3, \ldots)$. Evidently $A_i \subset R_i$ for each i. If $x \in X$, then $x \in R_i$ for some i. If i_x is the smallest such i, then

$$x \in A_{i_x}.$$

Hence $\mathfrak{A} = \{A_i\}$ is a refinement of $\mathfrak{R}$. Finally, observe that $R_i \cap A_j = \phi$ if $j > i$. Hence $\mathfrak{A}$ is locally finite.[4]

Now if X satisfies (i), then there is a locally finite collection $\{G_i\}$ of open sets such that $A_i \subset G_i$ for each i. Let $U_i = R_i \cap G_i$. If $x \in X$, then, for some i, $x \in A_i$. But $A_i \subset R_i$ and $A_i \subset G_i$. Hence $x \in R_i \cap G_i = U_i$. Therefore $\mathfrak{U} = \{U_i\}$ is an open refinement of $\mathfrak{R}$. Moreover, $\mathfrak{U}$ is locally finite: $N \cap U_i \neq \phi$ implies $N \cap G_i \neq \phi$. Hence $\mathfrak{U}$ is a locally finite open refinement of $\mathfrak{R}$. Therefore X is countably paracompact. This completes the proof of Theorem 3.

It is well known that if a topological space X satisfies condition (g) of Theorem 3, then X is normal (**10**, Chapter V, §§ 3.2, 4.1–4.5). Hence countably paracompact normal spaces are obtained as a generalization of fully normal spaces in the following natural way.

COROLLARY. *A topological space X is countably paracompact and normal if and only if each countable open covering of X admits a countable, open, star refinement:*

[4] This construction of $\mathfrak{A}$ is due to Michael (**3**, Lemma 2).

3. Remarks. The word "closed" in conditions (d), (e), and (f) of Theorem 3 cannot be replaced by "open": a countable open covering of an *arbitrary* space is a σ-discrete (and hence σ-locally finite and σ-closure-preserving) refinement of itself. Thus the validity of Theorem 3 with "closed" replaced by "open" in conditions (d), (e), and (f) would imply that every normal space is countably paracompact. Dowker **(1)** has given an example of a normal (albeit *not* Hausdorff) space which is not countably paracompact.

The word "closed" in condition (c) of Theorem 3 cannot be replaced by "open": the example due to Dowker referred to above has the property that *every* collection of open subsets is closure-preserving. It may well be, however, that a normal Hausdorff space is countably paracompact if and only if each countable open covering admits a (countable) closure-preserving open refinement.

The proof that (i) implies (a) in Theorem 3 shows that a countable open covering of an arbitrary space admits a locally finite refinement. Hence the following assertion is *false*: A normal space X is countably paracompact if and only if each countable open covering of X admits a locally finite refinement.

4. Some applications. Using conditions (b) and (d) of Theorem 3, it can easily be shown that every F_σ-subset of a countably paracompact normal space is countably paracompact (and, of course, normal **(8)**). The proof is analogous to that of **(3**, Proposition 3**)**. Similarly, conditions (b) and (d) can be employed to show that if a topological space X is the union of a locally finite collection of closed, normal, countably paracompact subspaces, then X is countably paracompact and normal. (This result is not new. See **(5**, Theorem 8.2 ff.**)**; observe that a locally finite closed covering of a space X dominates X.)

It follows from Theorem 3 (g) and a result due to Shirota **(7**, p. 23**)** that the family $\mathscr{C}$ of all countable open coverings of a countably paracompact normal space X is a base of a uniformity for X in the sense of Tukey **(10)**. In the Weil-Bourbaki version of uniformity **(2**, Chapter 6**)**, this means that the family $\{V(\mathfrak{R}) : \mathfrak{R} \in \mathscr{C}\}$, where $V(\mathfrak{R}) = \cup\{R \times R : R \in \mathfrak{R}\}$, is a base of a uniformity for X.

References

1. C. H. Dowker, *On countably paracompact spaces*, Can. J. Math., *3* (1951), 219–224.
2. J. L. Kelley, *General Topology* (New York, 1955).
3. E. Michael, *A note on paracompact spaces*, Proc. Amer. Math. Soc. , *4* (1953), 831–838.
4. ———, *Another note on paracompact spaces, to appear in Proc. Amer. Math. Soc.*
5. ———, *Continuous selections I*, Ann. of Math. (2), *63* (1956), 361–382.
6. K. Morita, *Star-finite coverings and the star-finite property*, Math. Japonicae, *1* (1948), 60–68.
7. T. Shirota, *A class of topological spaces*, Osaka Math. J., *4* (1952), 23–40.
8. Yu. Smirnov, *A necessary and sufficient condition for metrisability of a topological space*, Doklady Akad. Nauk SSSR. N.S., *77* (1951), 197–200.
9. A. H. Stone, *Paracompactness and product spaces*, Bull. Amer. Math. Soc., *54* (1948), 977–982.
10. J. Tukey, *Convergence and Uniformity in Topology* (Princeton, 1940).

Purdue University

GAMES ON A COMPACT SET

J. E. L. PECK and A. L. DULMAGE

1. Introduction. Von Neumann's fundamental theorem of the theory of games has been extended by various authors and recently in two different directions by Kneser **(8)** and Nikaidô **(9)**. We present here a form of the theorem, which is more general than that of both these authors. We develop some consequences of this theorem, which make it easier to decide whether certain classes of games have a value and we give several illustrative examples.

2. The fundamental theorem. Our theorem concerns the following situation: $\bar{X}$ is a convex subset of a real linear space, Y is any subset of a real linear space and $\bar{Y}$ is the convex set generated by Y. A real function f is defined on $\bar{X} \times \bar{Y}$ and $f(x,y)$ is a concave function of $x \in \bar{X}$ and a convex function of $y \in \bar{Y}$; i.e., for $0 \leqslant \lambda \leqslant 1$, we have both

$$f(\lambda x_1 + (1-\lambda)x_2, y) \geqslant \lambda f(x_1,y) + (1-\lambda) f(x_2,y)$$

and

$$f(x, \lambda y_1 + (1-\lambda)y_2) \leqslant \lambda f(x,y_1) + (1-\lambda) f(x,y_2).$$

THEOREM 1. *If $\bar{X}$ is compact in a topology* (no separation axiom is assumed) *which is such that for every $y \in Y$, $f(x,y)$ is an upper semi-continuous function of x, then*

$$\sup_{x \in \bar{X}} \inf_{y \in Y} f(x,y) \geqslant \inf_{y \in \bar{Y}} \sup_{x \in \bar{X}} f(x,y).$$

To prove this theorem we follow the procedure of Kneser **(8)** and establish the following

LEMMA. *If $\{f_1, \ldots, f_n\}$ is a finite set of real concave functions, which are upper semi-continuous on a compact, convex set K, and if for all $x \in K$, $\min_{1 \leqslant i \leqslant n} f_i(x) < 0$, then there is a mean*

$$g = \sum_{i=1}^{n} \lambda_i f_i, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad \lambda_i \geqslant 0, \qquad i = 1, \ldots, n$$

such that for all $x \in K$, $g(x) < 0$.

We prove the lemma first for two concave upper semi-continuous functions f and g defined on the compact convex set K, such that $\min(f(x),g(x)) < 0$. Let $F = \{x : f(x) \geqslant 0\}$ and $G = \{x : g(x) \geqslant 0\}$. These subsets are compact

Received October 11, 1956. This work was done at the Summer Research Institute of the Canadian Mathematical Congress.

and disjoint. If F, or G, is void, the proof is trivial. We thus assume that both F and G are not void.

From the hypothesis of upper semi-continuity we may find $p \in F$ and $q \in G$ such that

$$\max_{x \in F} f(x)/-g(x) = f(p)/-g(p) = \alpha \geqslant 0, \tag{1}$$

$$\max_{x \in G} g(x)/-f(x) = g(q)/-f(q) = \beta \geqslant 0. \tag{2}$$

Since $f(p) \geqslant 0$ and $f(q) < 0$, we may find $\rho > 0$ and $\sigma \geqslant 0$ so that $\rho + \sigma = 1$ and $\rho f(p) + \sigma f(q) = 0$. Because f is concave $f(\rho p + \sigma q) \geqslant 0$, whence by hypothesis and the concavity of g,

$$\rho g(p) + \sigma g(q) \leqslant g(\rho p + \sigma q) < 0.$$

Thus using (2) and (1),

$$\rho g(p) < -\sigma g(q) = \sigma\beta f(q) = -\rho\beta f(p) = \rho\alpha\beta g(p).$$

But $\rho > 0$ and $g(p) < 0$ so that $\alpha\beta < 1$. Now choose $\gamma > \alpha$, $\delta > \beta$ so that $\gamma\delta = 1$ and put

$$\lambda = (1 + \gamma)^{-1} = \delta(1 + \delta)^{-1}, \quad \mu = \gamma(1 + \gamma)^{-1} = (1 + \delta)^{-1}.$$

Then if $x \in F$, we have $f(x) \geqslant 0$, $g(x) < 0$ whence

$$\lambda f(x) + \mu g(x) \leqslant (\mu - \lambda\alpha)g(x) = (\gamma - \alpha)(1 + \gamma)^{-1}g(x) < 0.$$

Similarly, $\lambda f(x) + \mu g(x) < 0$ if $x \in G$. If $x \notin F$ and $x \notin G$, the statement is trivial.

To extend this to n functions $f_1, \ldots, f_n$, we put $F = \{x : f_n(x) \geqslant 0\}$. The set F is compact and convex. Supposing the theorem true for the functions $f_1, \ldots, f_{n-1}$ on the set F, we have

$$h = \mu_1 f_1 + \ldots + \mu_{n-1} f_{n-1}, \quad \sum_{i=1}^{n-1} \mu_i = 1, \ \mu_i \geqslant 0, \quad i = 1, \ldots, n-1,$$

such that for all $x \in F$, $h(x) < 0$. Since a mean of concave upper semi-continuous functions is concave and upper semi-continuous, we may apply the case $n = 2$ to the functions f_n and h on K to obtain $g = \lambda h + \mu f_n$ such that for all $x \in K$, $g(x) < 0$. Putting

$$\lambda_i = \lambda\mu_i, \qquad i = 1, \ldots, n-1$$

and $\lambda_n = \mu$ we have

$$g = \sum_{i=1}^{n} \lambda_i f_i, \quad \sum_{i=1}^{n} \lambda_i = 1, \ \lambda_i \geqslant 0, \qquad i = 1, \ldots, n.$$

This completes the proof of the lemma.

To prove Theorem 1, put

$$v = \inf_{y \in Y} \sup_{x \in X} f(x,y). \tag{3}$$

We shall establish that

$$\exists_{x \in \hat{X}} \forall_{y \in Y},\ f(x,y) \geqslant v. \tag{4}$$

Suppose (4) is not true, then

$$\forall_{x \in \hat{X}} \exists_{y \in Y},\ f(x,y) < v. \tag{5}$$

Put $G_y = \{x : f(x,y) < v\}$, then by (5) and the hypothesis, $\{G_y : y \in Y\}$ is an open covering of $\hat{X}$. Because $\hat{X}$ is compact, we may select $y_1, \ldots, y_n \in Y$ so that

$$\hat{X} = G_{y_1} \cup \ldots \cup G_{y_n}.$$

Thus

$$\forall_{x \in \hat{X}} \exists_{i \in \{1, \ldots, n\}},\ f(x,y_i) < v,$$

so that the functions $f_i(x) = f(x,y_i) - v$, $i = 1, \ldots, n$ satisfy the hypothesis of the lemma. For some

$$y = \sum_{i=1}^{n} \lambda_i y_i \in \hat{Y},$$

we therefore have

$$\forall_{x \in \hat{X}},\ f(x,y) < \sum_{i=1}^{n} \lambda_i f(x,y_i) < v.$$

However $\sum \lambda_i f(x,y_i)$, being an upper semi-continuous function of $x \in \hat{X}$, achieves its maximum. This is a contradiction of (3) so that (4) is true. But (4) means that

$$\sup_{x \in \hat{X}} \inf_{y \in Y} f(x,y) \geqslant v = \inf_{y \in \hat{Y}} \sup_{x \in \hat{X}} f(x,y).$$

We have immediately two corollaries. The first is a consequence of the well-known inequality

$$\sup_{x \in \hat{X}} \inf_{y \in \hat{Y}} f(x,y) \leqslant \inf_{y \in \hat{Y}} \sup_{x \in \hat{X}} f(x,y), \tag{6}$$

and the second is a consequence of (6) and the fact that if $f(x,y)$ is linear in $y \in \hat{Y}$, then

$$\inf_{y \in Y} f(x,y) = \inf_{y \in \hat{Y}} f(x,y).$$

COROLLARY 1. *If* $\hat{Y} = Y$, *then*

$$\sup_{x \in \hat{X}} \inf_{y \in \hat{Y}} f(x,y) = \inf_{y \in \hat{Y}} \sup_{x \in \hat{X}} f(x,y).$$

COROLLARY 2. *If* $f(x,y)$ *is linear in* $y \in \hat{Y}$, *then*

$$\sup_{x \in \hat{X}} \inf_{y \in \hat{Y}} f(x,y) = \inf_{y \in \hat{Y}} \sup_{x \in \hat{X}} f(x,y).$$

The first corollary generalizes both the theorems of Kneser **(8)** and Nikaidô **(9)**. Our referee points out that it is the theorem of Pettis **(10**, Theorem 5**)** and was observed by Kuhn in his review of Kneser's paper for Mathematical Reviews **(14**, 301**)**. In the following section we exploit Corollary 2.

We show now that both the above corollaries generalize a theorem of Karlin **(7**, Theorem 3**)**. We write $f_x(y) = f(x,y)$ and $F = \{f_x : x \in \hat{X}\}$. We define a mapping T from $\hat{X}$ onto F by $Tx = f_x$, and we put $(y,Tx) = f(x,y)$. We introduce on F a topology $\mathfrak{F}$ which is generated by the sub-base

$$\{\{Tx : (y,Tx) < a\} : y \in Y, a \in \text{reals}\}.$$

We shall say that the mapping T is *weakly compact* if the range F of T is compact in the topology $\mathfrak{F}$. Since

$$\{Tx : (y,Tx) < a\} = \{f_x : f(x,y) < a\},$$

it follows that the set F is $\mathfrak{F}$-compact if and only if the convex set $\hat{X}$ is compact in the topology $\mathfrak{X}$ generated by the sub-base

$$\{\{x : f(x,y) < a\} : y \in Y, a \in \text{reals}\}.$$

We thus have that $\hat{X}$ is $\mathfrak{X}$-compact if and only if T is a weakly compact mapping. However $\mathfrak{X}$ is a topology for Theorem 1, so that *the hypothesis of Theorem* 1 *is equivalent to the statement that* T *is weakly compact.*

In this form we may see that Theorem 1 and its corollaries generalize the theorem of Karlin cited above, for Karlin assumes that $\hat{Y} = Y$, that $f(x,y)$ is linear in y and that T is weakly compact using a topology which is finer than $\mathfrak{F}$ and is therefore weakly compact in the above sense.

The next theorem shows that it is possible to determine whether a game has a value, by examining another associated game.

THEOREM 2. *If* f *is a function defined on* $\hat{X} \times Y$ *and if for each* $y \in Y$ *there exist numbers* $\phi(y) > 0$ *and* $\psi(y)$ *such that* $\hat{X}$ *is compact in a topology in which for all* $y \in Y$, $g(x,y) = \phi(y)f(x,y) + \psi(y)$ *is a concave upper semi-continuous function of* x, *then*

$$\sup_{x \in \hat{X}} \inf_{y \in \hat{Y}} f(x,y) = \inf_{y \in \hat{Y}} \sup_{x \in \hat{X}} f(x,y),$$

where f *is extended to* $\hat{X} \times \hat{Y}$ *by linearity in* y.

Since the set

$$\{x : g(x,y) < a\} = \{x : f(x,y) < -\psi(y) + a/\phi(y)\},$$

$\hat{X}$ is also compact in a topology such that for each $y \in Y$, $f(x,y)$ is an upper semi-continuous function of x. The theorem therefore follows by Corollary 2.

If the function f is such that for each $y \in Y$,

$$(7) \qquad -\infty < \inf_{x \in \hat{X}} f(x,y) < \sup_{x \in \hat{X}} f(x,y) < +\infty,$$

then $\phi(y)$ and $\psi(y)$ can be chosen so that for each $y \in Y$,

$$\sup_{x \in \bar{X}} g(x,y) = 1 \text{ and } \inf_{x \in \bar{X}} g(x,y) = 0.$$

Thus a game $(\bar{X}, Y, f)$ satisfying (7) has a value by Corollary 2, if and only if a certain associated normalized game $(\bar{X}, Y, g)$ has a value by Corollary 2.

3. Special cases. If the hypothesis of Theorem 1 is restricted further, we obtain the following, which is a generalization of the theorems of Choquet **(3)**, Ville **(11)** and Wald **(12)**.

THEOREM 3. *If X is a compact Hausdorff space, Y is any set and f is a real function on $X \times Y$, which is such that for each y, $f(x,y)$ is a continuous function of x, then*

$$\sup_\xi \inf_\eta f(\xi, \eta) = \inf_\eta \sup_\xi f(\xi, \eta)$$

where ξ ranges over the normalized measures on the Borel subsets of X, η ranges over the finite mixtures

$$\eta = \sum_{k=1}^{n} \eta_k y_k, \quad \sum_{k=1}^{n} \eta_k = 1, \eta_k \geqslant 0, \qquad k = 1, \ldots, n$$

and

$$f(\xi,\eta) = \sum_{k=1}^{n} \eta_k \int_X f(x,y_k) d\xi(x).$$

This is a consequence of Theorem 1. Let $C(X)$ be the Banach space of all continuous real functions on X, and $C^*(X)$ its conjugate space. The space $\bar{X}$ of normalized measures on the Borel subsets of X is a closed convex subset **(6)** of the weak* compact unit sphere in $C^*(X)$. This is a topology such that if $\phi \in C(X)$ then $\phi(\xi) = \int \phi d\xi$ is a continuous function of ξ, so it follows that $f(\xi,y) = \int_X f(x,y) d\xi(x)$ is a continuous function of ξ. The hypotheses of Theorem 1 and its Corollary 2 are therefore satisfied.

The same theorem may be rephrased in the following form.

THEOREM 3′. *If F is a subset of the space $C(X)$ of real continuous functions on a compact Hausdorff space, if $\bar{F}$ is the convex set generated by F and $\bar{X}$ is the set or normalized measures on the Borel subsets of X then*

$$\sup_{\xi \in \bar{X}} \inf_{f \in F} \int_X f d\xi = \inf_{f \in \bar{F}} \sup_{x \in X} f(x).$$

We may weaken the hypothesis of the compactness of X if we restrict the pay-off function. Thus we obtain

THEOREM 4. *If X is a locally compact Hausdorff space, if Y is any set and if f is a function on $X \times Y$ such that for each $y \in Y$, there is a number $m(y)$ such that $f(x,y) - m(y)$ is a non-negative continuous function which vanishes at infinity, then*

$$\sup_\xi \inf_\eta f(\xi, \eta) = \inf_\xi \sup_\eta f(\xi, \eta)$$

where ξ ranges over the normalized measures on the Borel subsets **(5)** *of X, η ranges over the finite mixtures*

$$\eta = \sum_{k=1}^{n} \eta_k y_k, \quad \sum_{k=1}^{n} \eta_k = 1, \eta_k \geqslant 0, \qquad k = 1, \ldots, n$$

and

$$f(\xi,\eta) = \sum_{k=1}^{n} \eta_k \int_X f(x,y_k)d\xi(x).$$

To prove this theorem we let $X_\infty = X \cup \{\infty\}$ be the one-point compactification of X and we define $f(\infty,y) = m(y)$. The hypothesis of Theorem 3 is therefore satisfied for f on $X_\infty \times Y$. We thus have

$$\sup_{\xi \in \bar{X}_\infty} \inf_{\eta \in Y} f(\xi,\eta) = \inf_{\eta \in Y} \sup_{\xi \in \bar{X}_\infty} f(\xi,\eta), \tag{8}$$

where $\bar{X}_\infty$ is the set of normalized measures on the Borel subsets of X_∞ and $f(\xi,\eta)$ is defined as above with the integral taken over X_∞. Let $\bar{X}$ be the set of normalized measures on the Borel subsets of X. We shall prove that

$$\sup_{\xi \in \bar{X}} \inf_{\eta} f(\xi,\eta) \geqslant \sup_{\xi \in \bar{X}_\infty} \inf_{\eta} f(\xi,\eta), \tag{9}$$

from which the result follows by using (8) and (6).

Let $\zeta \in \bar{X}_\infty$ be chosen arbitrarily and suppose that $\zeta(X) = \theta$. Choose any $\xi' \in \bar{X}$ and define ξ by $\xi(E) = (1-\theta)\xi'(E) + \zeta(E)$ for every Borel subset E of X. Then because $\theta \leqslant 1$ we have that $\xi \in \bar{X}$, and that for all E, $\zeta(E) \leqslant \xi(E)$. Now put

$$g(x,y) = f(x,y) - m(y).$$

Since $g(x,y) \geqslant 0$ and $g(\infty,y) = 0$, we have that

$$\forall_{y \in Y} \int_X g(x,y)d\xi(x) \geqslant \int_X g(x,y)d\zeta(x) = \int_{X_\infty} g(x,y)d\zeta(x).$$

Thus

$$\forall_{\zeta \in \bar{X}_\infty} \exists_{\xi \in \bar{X}} \forall_{y \in Y}, \ f(\xi,y) \geqslant f(\zeta,y),$$

and therefore

$$\forall_{\zeta \in \bar{X}_\infty} \exists_{\xi \in \bar{X}}, \inf_{\eta \in Y} f(\xi,\eta) \geqslant \inf_{\eta \in Y} f(\zeta,\eta),$$

whence we may obtain the statement (9).

If in the above we suppose, for simplicity, that for all y, $m(y) = 0$, we have the following analogue of Theorem 3′:

THEOREM 4′. *If $C(X)$ is the space of continuous real functions which vanish at infinity on a locally compact Hausdorff space X, if F is a subset of $C(X)$ consisting of non-negative functions, and $\bar{F}$ is the convex set generated by F, then*

$$\sup_{\xi \in \bar{X}} \inf_{f \in F} \int_X f d\xi = \inf_{f \in \bar{F}} \sup_{x \in X} f(x).$$

We note that in Theorem 4, if X is a set with its discrete topology, then we obtain the theorem of Dulmage and Peck **(4)**.

4. Examples. We now give some examples to illustrate the results obtained above. The first example shows the usefulness of Theorem 1, Corollary 2, in determining whether a game has a value. What is significant is that, in investigating the compactness of the space of mixed strategies $\bar{X}$, it is not necessary to consider the upper semi-continuity of $f(x,y)$ for *all* mixed strategies y, but only for pure strategies $y \in Y$, This means that we are examining $\bar{X}$ in a weaker topology. Indeed it may happen that $\bar{X}$ is compact in this weaker topology and not compact in the topologies considered by Kneser **(8)** and Nikaidô **(9)**. The following example illustrates this situation.

Example 1. Two persons each choose a natural number, say m and n. If $m = n$ the pay-off is 2^m, otherwise it is zero. The spaces of pure strategies X and Y are thus the natural numbers and the spaces of mixed strategies $\bar{X}$ and $\bar{Y}$ are probability distributions on X and Y respectively. The pay-off function is $f(m,n) = 2^m\delta_{m,n}$, which when extended to mixed strategies is

$$f(x,y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} 2^m \delta_{m,n} x_m y_n = \sum_{i=1}^{\infty} 2^i x_i y_i.$$

We show that $\bar{X}$ is not compact in the weakest topology such that for all $y \in \bar{Y}$, $f(x,y)$ is an upper semi-continuous function of x.

Let $y^{(n)} = \{0, 0, \ldots, 0, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\} \in \bar{Y}$ where the weight on the first n coordinates is zero. Then for each $x \in \bar{X}$ we have

$$f(x,y^{(n)}) = 2^n(x_{n+1} + x_{n+2} + x_{n+3} + \ldots).$$

Let $G_n = \{x: f(x,y^{(n)}) < 2^n\}$, then each G_n is an open set in the topology considered. But since $G_n = \{x: x_{n+1} + x_{n+2} + \ldots < 1\}$ the family $\{G_n: n = 1, 2, \ldots\}$ is an open covering of $\bar{X}$ which has no finite sub-covering.

However, we must also show that $\bar{Y}$ is not compact in the weakest topology such that for all $x \in \bar{X}$, $f(x,y)$ is a lower semi-continuous function of y. But this is simple, since if

$$H_m = \{y : f(m,y) > 0\} = \{y : y_m > 0\},$$

then $\{H_m : m = 1, 2, \ldots\}$ is an open covering with no finite sub-covering.

The compactness of $\bar{X}$ (and of $\bar{Y}$) thus fails in the stronger topology used by Kneser and Nikaido. However it is not difficult to show that $\bar{X}$ is compact in the topology of Theorem 1. As a matter of fact, this game has a value, also because of Theorem 4, or the theorem in **(4)**. The value of the game is one. *Example* 1 does not satisfy the hypothesis of the theorem of Blackwell and Girshick **(2**, Theorem 2.3.3**)**.

A more interesting example, similar to the above, which does not satisfy the hypothesis in **(4)**, but does have a value by Theorem 4 is the following.

Example 2. Two players each choose a real number, say x and y, and the pay-off is $f(x,y) = \psi(x - y)e^y$, where $\psi(s) = 1 - |s|$ for $|s| \leqslant 1$ and $\psi(s) = 0$ for $|s| > 1$. The value of the game is zero. If the game is played over the quadrant $x \geqslant a$, $y \geqslant a$, its value is positive and lies between $\frac{1}{2}e^{a-1}$ and e^{a+1}.

The third and fourth are examples of simple games, both of which satisfy the conditions of Theorem 1, (both have the value zero) but which do not satisfy the hypotheses of the theorems of Karlin **(7**, Theorem 3 and remark 4**)**, Berge **(1**, § 25**)**, and Wald **(13)**.

Example 3. Both X and Y are the closed interval $[0,1]$ and $f(x,y) = 1$ if $x = y$, $f(x,y) = 0$ if $x \neq y$.

Example 4. Both X and Y are the sets of natural numbers and the pay-off function is the same as in Example 3.

Example 4 is the associated normalized form of Example 1 in the sense of Theorem 2.

The fifth is an example which shows that Theorem 1 still has its limitations. This is a game which does not satisfy the hypothesis of Theorem 1, but satisfies the hypotheses of the theorems of Blackwell and Girschik **(2**, Theorem 2.3.3**)**, and Wald **(13)**. Moreover the operator T is completely continuous in the sense of Karlin **(7**, remark 4**)**.

Example 5. The sets of pure strategies X and Y are the natural numbers and the pay-off function is $f(m,n) = 1$ for $m \geqslant n$, and $f(m,n) = 1 - 2^{-n}$ for $m < n$. We show that neither $\hat{X}$ nor $\hat{Y}$ are compact in the topology of Theorem 1. Let

$$G_n = \{x : f(x,n) < 1 - 2^{-n}\}.$$

Given $x \in \hat{X}$ we may choose n so that $\sum_{m>n} x_m < 2^{-n}$, whence

$$f(x,n) = \tfrac{1}{2}x_1 + \tfrac{3}{4}x_2 + \ldots + (1 - 2^{1-n})x_{n-1} + \sum_{m>n} x_m < (1 - 2^{1-n}) + 2^{-n} = 1 - 2^{-n}.$$

Therefore $\{G_n : n = 1, 2, \ldots\}$ is an open covering of $\hat{X}$. However if x is the nth pure strategy $x = n$, we have for $j \leqslant n$ that $f(x,j) = 1 > 1 - 2^{-j}$, so that $x \notin G_1 \cup \ldots \cup G_n$. Thus $\hat{X}$ is not compact.

Now put $H_m = \{y : f(m,y) > 1 - 2^{-m}\}$. Given $y \in \hat{Y}$ we choose m so that y_m is the first non-zero coefficient of y. Then

$$f(m,y) = y_m + (1 - 2^{-m})(y_{m+1} + y_{m+2} + \ldots) > 1 - 2^{-m},$$

so that $y \in H_m$. Therefore $\{H_m : m = 1, 2, \ldots\}$ is an open covering of $\hat{Y}$. However it has no finite sub-covering because if y is the $(m + 1)$th pure strategy $y = m + 1$ we have for $i \leqslant m$, $f(i,m) = 1 - 2^{-i}$, so that $y \notin H_i$. Thus $\hat{Y}$ is also not compact.

We shall omit verification that the game satisfies the hypotheses of the theorems cited above.

References

1. C. Berge, *Sur une théorie ensembliste des jeux alternatifs*, J. Math. Pure Appl., (9), *32* (1953), 129–184.
2. D. Blackwell and M. A. Girschick, *Theory of games and statistical decisions* (New York, 1954).
3. A Choquet, *Sur le théorème des point-selle de la théorie des jeux*, Bull. Sci. Math., (2), *79* (1955), 48–53.
4. A. L. Dulmage and J. E. L. Peck, *Certain infinite zero-sum two-person games*, Can. J. Math., *8* (1956), 412–416.
5. P. R. Halmos, *Measure Theory* (New York, 1950).
6. S. Kakutani, *Concrete representations of abstract (M)-spaces*, Ann. of Math., *42* (1941), 994–1024.
7. S. Karlin, *Operator treatment of the minimax principle*, Contributions to the theory of games, Ann. of Math. Studies, No. 24 (1950), 133–154.
8. H. Kneser, *Sur un théorème fondamental de la théorie des jeux*, C.R. Acad. Sci. Paris, *234* (1952), 2418–2420.
9. H. Nikaidô, *On a minimax theorem and its applications to functional analysis*, J. Math. Soc. Japan, *5* (1953), 86–94.
10. B. J. Pettis, *Separation theorems for convex sets*, Math. Mag. *29* (1956), 233–245.
11. J. Ville, *Sur la théorie générale des jeux ou intervient l'habilité des joueurs*, Traité du Calcul des Probabilités et de ses Applications, E. Borel, etc., *4*, *2* (1938), 105–113.
12. A. Wald, *Generalization of a theorem by von Neumann concerning zero sum two person games*, Ann. of Math. *46* (1945), 281–286.
13. A. Wald, *Foundations of a general theory of sequential decision functions*, Econometrica, *15* (1947), 279–313.

McGill University *Royal Military College of Canada*

ON THE INVERSION OF THE GAUSS TRANSFORMATION

P. G. ROONEY

1. Introduction. The inversion theory of the Gauss transformation has been the subject of recent work by several authors. If the transformation is defined by

$$f(x) = \mathscr{G}(\phi(x)) = (4\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x-t)^2} \phi(t)dt, \tag{1.1}$$

then operational methods indicate that

$$\exp(-D^2) f(x) = \phi(x),$$

under a suitable definition of the differential operator. Hirschmann and Widder **(5**; chap. 8**)** have shown that this is indeed the case if the operator is defined by use of the formula

$$e^{-y^2} = \lim_{t \to 1-} \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} K(s,t) e^{sy} ds,$$

where

$$K(s,t) = (\pi/t)^{\frac{1}{2}} \exp(s^2/4t).$$

Also Widder **(8**; § 1**)** makes use of the formula

$$e^{-y^2} = \pi^{-\frac{1}{2}} \int_0^{\infty} e^{-\frac{1}{4}t^2} \cos yt dt$$

to define $\exp(-D^2)$.

A more straightforward approach seems to be to use the power series for

$$e^{-y^2}$$

This we shall do here, defining

$$e^{-D^2} f(x) = \sum_{n=0}^{\infty} (-1)^n f^{(2n)}(x)/n! \tag{1.2}$$

This approach seems to have been originally suggested by Eddington **(2)**. Eddington did not study the convergence of this series other than to note that for some very simple Gauss transforms, the series diverges. Widder **(8**; § 1**)** remarks that for some very simple transforms the series is not even summable in the Abel sense. Pollard **(6**; § 8**)** showed that if $\phi \in L_2(-\infty, \infty)$, the series converges in the L_2-sense to ϕ, that is

Received September 13, 1956; in expanded form October 3, 1956.

$$\phi(x) = \text{l.i.m.}_{r\to\infty} \sum_{n=0}^{r} (-1)^n f^{(2n)}(x)/n!\,,$$

but nothing seems to be known of the pointwise convergence of the series.

Here we propose to give sufficient conditions for the series to converge and to be Abel summable to $\phi(x)$. This occupies §§ two and three respectively. In § four we make some applications of our results to the study of Hermite polynomials.

2. Convergence of the series. Theorem 1 gives sufficient conditions for the series II to converge at a point to the mean value of ϕ there.

THEOREM 1. *If*

(i) $\quad \phi \in L(-\delta, \delta) \quad \delta > 0$, and $|t|^{\lambda} e^{-(x_0-t)^2/8}\phi(t) \in L(-\infty, \infty)$,

for some $\lambda > 3$,

(ii) ϕ *is of bounded variation in a neighbourhood of* x_0, *then* $\mathscr{G}(\phi(x))$ *exists for all* x, *and the series* 1.2 *converges at* $x = x_0$ *to* $\frac{1}{2}\{\phi(x_0+) + \phi(x_0-)\}$.

Proof. The existence of $\mathscr{G}(\phi(x))$ is clear. Now by **(3**; 10.13(7) and 10.13(2)**)**,

$$D^{2r} e^{-\frac{1}{2}x^2} = 2^{-2r} e^{-\frac{1}{2}x^2} H_{2r}(\tfrac{1}{2}x) = e^{-\frac{1}{2}x^2} r!(-1)^r L_r^{-\frac{1}{2}}(\tfrac{1}{4}x^2).$$

But since $f(x)$ is a bilateral Laplace transform, it follows from **(1**; chap. 3 § 2**)** that we may differentiate the integral for f under the integral sign as often as we choose. Thus

$$f^{(2r)}(x) = \frac{r!(-1)^r}{(4\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x-t)^2} L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x-t)^2)\phi(t)dt.$$

Hence if $s_n(x)$ denotes the nth partial sum of the series 1.2, we have, using **(3**; 10.12(38)**)**,

$$\begin{aligned} s_n(x_0) &= \sum_{r=0}^{n} (-1)^r f^{(2r)}(x_0)/r! \\ &= \frac{1}{(4\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x_0-t)^2} \sum_{r=0}^{n} L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x_0-t)^2)\phi(t)dt \\ &= \left(\int_{-\infty}^{x_0 - n^{1/2\lambda}} + \int_{x_0 - n^{1/2\lambda}}^{x_0 + n^{1/2\lambda}} + \int_{x_0 + n^{1/2\lambda}}^{\infty} \right) L_n^{\frac{1}{2}}(\tfrac{1}{4}(x_0-t)^2)\phi(t)dt \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Consider I_3. By **(3**; 10.18(14)**)**,

$$e^{-\frac{1}{2}x}|L_n^{\alpha}(x)| < \Gamma(\alpha+1+n)/n!\,\Gamma(\alpha+1), \qquad \alpha > 0.$$

Hence

$$\begin{aligned} |I_3| &< \frac{\Gamma(n+3/2)}{\Gamma(3/2)n!(4\pi)^{\frac{1}{2}}} \int_{x_0 + n^{1/2\lambda}}^{\infty} e^{-(x_0-t)^2/8} |\phi(t)|dt \\ &< A\,\frac{\Gamma(n+3/2)}{n!(x_0 + n^{1/2\lambda})^{\lambda}} \int_{x_0 + n^{1/2\lambda}}^{\infty} t^{\lambda} e^{-(x_0-t)^2/8} |\phi(t)|dt. \end{aligned}$$

But from Stirling's formula, $\Gamma(n + 3/2)/n! = O(n^{\frac{1}{2}})$ as $n \to \infty$, and thus

$$\Gamma(n + 3/2)/n!(x_0 + n^{\frac{1}{2}\lambda})^\lambda = O(1) \quad \text{as} \quad n \to \infty,$$

so that $I_3 = o(1)$ as $n \to \infty$. Similarly $I_1 = o(1)$ as $n \to \infty$.

Now by (3; 10.15(2)) and using Stirling's formula, since $\lambda > 3$,

$$L_n^{\frac{1}{2}}(\tfrac{1}{4}x^2) \sim (2/\pi)^{\frac{1}{2}} e^{x^2/8} \sin\left((n + 3/4)^{\frac{1}{2}}x\right)/x$$

as $n \to \infty$, uniformly in $-n^{1/2\lambda} \leqslant x \leqslant n^{1/2\lambda}$. Hence

$$I_2 \sim \frac{1}{\pi} \int_{x_0 - n^{1/2\lambda}}^{x_0 + n^{1/2\lambda}} e^{-(x_0 - t)^2/8} \frac{\sin\left((n + 3/4)^{\frac{1}{2}}(x_0 - t)\right)}{x_0 - t} \phi(t)dt$$

$$= \int_{-\infty}^{\infty} - \int_{-\infty}^{x_0 - n^{1/2\lambda}} - \int_{x_0 + n^{1/2\lambda}}^{\infty} = I_4 + I_5 + I_6.$$

For I_6 we have if $n > 0$,

$$|I_6| \leqslant \int_{x_0 + n^{1/2\lambda}}^{\infty} e^{-(x_0 - t)^2/8} \phi(t)/(t - x_0)dt,$$

so that $I_6 = o(1)$ as $n \to \infty$, and similarly $I_5 = o(1)$ as $n \to \infty$.

But by the Fourier integral theorem,

$$I_4 \to \tfrac{1}{2}\{\phi(x_0 +) + \phi(x_0 -)\},$$

so that

$$\lim_{n \to \infty} s_n(x_0) = \tfrac{1}{2}\{\phi(x_0 +) + \phi(x_0 -)\},$$

and the theorem is proved.

3. Abel summability of the series.

THEOREM 2. *If*

1. $e^{-(x_0 - t)^2/8}\phi(t) \in L(-\infty, \infty)$,
2. $\phi(x_0 +)$ and $\phi(x_0 -)$ exist,

then $\mathscr{G}(\phi(x))$ *exists for all* x, *and the series* II *is Abel summable for* $x = x_0$ *to* $\tfrac{1}{2}\{\phi(x_0 +) + \phi(x_0 -)\}$.

Proof. The existence of $\mathscr{G}(\phi(x))$ is clear. If

$$u(x,t) = \sum_{r=0}^{\infty} (-t)^r f^{(2r)}(x)/r!, \qquad 0 \leqslant t < 1,$$

we must show that $u(x_0, 1 -) = \tfrac{1}{2}\{\phi(x_0 +) + \phi(x_0 -)\}$. But as in the proof of Theorem 1,

$$f^{(2r)}(x) = \frac{r!(-1)^r}{(4\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x-y)^2} L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x - y)^2)\phi(y)dy.$$

so that using (3; 10.12(17)), if $0 \leqslant t < 1$,

$$u(x_0,t) = (4\pi)^{-\frac{1}{2}} \sum_{r=0}^{\infty} t^r \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x_0-y)^2} L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x_0-y)^2)\phi(y)dy$$

$$= \frac{1}{(4\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x_0-y)^2} \sum_{r=0}^{\infty} t^r L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x_0-y)^2)\phi(y)dy$$

$$= \frac{1}{(4\pi(1-t))^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-(x_0-y)^2/4(1-t)}\phi(y)dy,$$

provided we justify the interchange of summation and integration.

For this it suffices to show that

$$K(x_0,t) = (4\pi)^{-\frac{1}{2}} \sum_{r=0}^{\infty} t^r \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x_0-y)^2} |L_r^{-\frac{1}{2}}(\tfrac{1}{4}(x_0-y)^2)\phi(y)|dy$$

is finite. But from (3; 10.18(15)) it follows that

$$|L_n^{-\frac{1}{2}}(x)| \leqslant 2e^{\frac{1}{2}x},$$

so that

$$K(x_0,t) \leqslant \frac{1}{\pi^{\frac{1}{2}}(1-t)} \int_{-\infty}^{\infty} e^{-(x_0-y)^2/8}|\phi(y)|dy < \infty$$

for $0 \leqslant t < 1$, and the interchange is justified.

Hence if $0 \leqslant t < 1$, $\delta > 0$,

$$u(x_0,t) = \frac{1}{(4\pi(1-t))^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-(x_0-y)^2/4(1-t)}\phi(y)dy$$

$$= \int_{-\infty}^{x_0-\delta} + \int_{x_0-\delta}^{x_0+\delta} + \int_{x_0+\delta}^{\infty} = I_1 + I_2 + I_3.$$

Consider I_3. Since $e^{-(x_0-y)^2 t/4(1-t)}$ decreases for $y > x_0$, we have

$$|I_3| \leqslant \frac{1}{(4\pi(1-t))^{\frac{1}{2}}} \int_{x_0+\delta}^{\infty} e^{-(x_0-y)^2/4(1-t)}|\phi(y)|dy$$

$$= \frac{1}{(4\pi(1-t))^{\frac{1}{2}}} \int_{x_0+\delta}^{\infty} e^{-(x_0-y)^2 t/4(1-t)} e^{-\frac{1}{4}(x_0-y)^2}|\phi(y)|dy$$

$$\leqslant \frac{e^{-\delta^2 t/4(1-t)}}{(4\pi(1-t))^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{4}(x_0-y)^2}|\phi(y)|dy \to 0 \quad \text{as } t \to 1-.$$

Similarly $I_1 \to 0$ as $t \to 1-$.

Finally from (7; chap. VII, Theorem 2b and corollaries 2b.2 and 2b.3), $I_2 \to \frac{1}{2}\{\phi(x_0+) + \phi(x_0-)\}$.

COROLLARY. *If*

$$e^{-(x_0-t)^2/8}\phi(t) \in L(-\infty, \infty)$$

and x_0 is in the Lebesgue set of ϕ, then $\mathcal{G}(\phi(x))$ exists for all x, and the series II *is Abel summable for $x = x_0$ to $\phi(x_0)$.*

The proof is identical with the proof of the theorem except that we do not use (7; Chap. VII, Corollary 2b.3).

It may be mentioned that

$$\frac{e^{t^2/8}}{1+t^2}$$

provides, with $x_0 = 0$, an example of a function satisfying the hypotheses of Theorem 2, but not those of Theorem 1.

4. Applications. Theorem 1 yields various results about Hermite polynomials. For example, it follows from **(3; 10.13(30))** that

$$\mathcal{G}\ (H_n(ax)) = (1-4a^2)^{\frac{1}{2}n}H_n(ax/(1-4a^2)^{\frac{1}{2}}), \qquad -\tfrac{1}{2} < a < \tfrac{1}{2}.$$

Now from **(3; 101.3(13))**,

$$D^rH_n(bx) = \begin{cases} (2b)^r \cdot n(n-1)\dots(n-r+1)H_{n-r}(bx), & r \leqslant n \\ 0, & r > n. \end{cases}$$

Hence from Theorem 1 we have, if $-\frac{1}{2} < a < \frac{1}{2}$,

$$H_n(ax) = (1-4a^2)^{\frac{1}{2}n}\sum_{r=0}^{[\frac{1}{2}n]} \frac{n(n-1)\dots(n-2r+1)}{r!} \cdot(-4a^2/(1-4a^2))^rH_{n-2r}(ax/(1-4a^2)^{\frac{1}{2}}),$$

or changing $ax/(1-4a^2)^{\frac{1}{2}}$ to x, and letting $\lambda = (1-4a^2)^{\frac{1}{2}}$,

$$H_n(\lambda x) = \sum_{r=0}^{[\frac{1}{2}n]} \frac{n(n-1)\dots(n-2r+1)}{r!}(\lambda^2-1)^r\lambda^{n-2r}H_{n-2r}(x),$$

the result being clearly valid now for all λ.

Another interesting application comes about as follows. An easy calculation shows that for all real v,

$$\mathcal{G}\ ((1+4v^2)^{-\frac{1}{2}}e^{v^2x^2/(1+4v^2)}) = e^{v^2x^2} = f(x),$$

and it follows easily from differentiating this formula that

$$\mathcal{G}\ ((1+4v^2)^{-3/2}xe^{v^2x^2/(1+4v^2)}) = xe^{v^2x^2} = g(x).$$

Theorem 1 can be applied to these functions if $-\frac{1}{2} < v < \frac{1}{2}$. Hence since, from **(3; (10.13(7))**

$$f^{(2r)}(x) = D^{2r}e^{v^2x^2} = (-1)^rv^{2r}e^{v^2x^2}H_{2r}(ivx),$$

and using **(3; 10.13(10))** and Leibnitz's rule,

$$\begin{aligned} g^{(2r)}(x) &= D^{2r}xe^{v^2x^2} = xD^{2r}e^{v^2x^2} + 2rD^{2r-1}e^{v^2x^2} \\ &= (-1)^rv^{2r-1}e^{v^2x^2}(2ivxH_{2r}(ivx) - 4rH_{2r-1}(ivx))/2i \\ &= (-1)^rv^{2r-1}e^{v^2x^2}H_{2r+1}(ivx)/2i. \end{aligned}$$

Thus from Theorem 1, if $-\frac{1}{2} < v < \frac{1}{2}$,

$$(1+4v^2)^{-\frac{1}{2}}\, e^{v^2x^2/(1+4v^2)} = e^{v^2x^2} \sum_{r=0}^{\infty} v^{2r} H_{2r}(ivx)/r!,$$

and

$$(1+4v^2)^{-3/2} x\, e^{v^2x^2/(1+4v^2)} = (2iv^2)^{-1}\, e^{v^2x^2} \sum_{r=0}^{\infty} v^{2r+1} H_{2r+1}(ivx)/r!,$$

whence on dividing we obtain

$$(1+4v^2)^{-\frac{1}{2}}\, e^{4v^4x^2/(1+4v^2)} = \sum_{r=0}^{\infty} v^{2r} H_{2r}(ivx)/r!$$

and

$$(1+4v^2)^{-3/2} \cdot 2iv^2 x\, e^{4v^4x^2/(1+4v^2)} = \sum_{r=0}^{\infty} v^{2r+1} H_{2r+1}(ivx)/r!$$

Now for every v, $-\frac{1}{2} < v < \frac{1}{2}$, each series of the last pair converges uniformly with respect to x for x in any closed bounded region of the complex x-plane. This follows for the first series, for example from the M–test on making use of the inequality

$$|H_{2r}(x)| \leqslant 2^{2r+1}\, r!\, e^{\frac{1}{2}x^2}$$

given in **(3; 10.18(16))**, and for the second series similarly. Hence by the principal of analytic continuation, the series converge to the indicated sums for all complex x, and another appeal to analytic continuation with respect to v shows that the formulae hold for all complex v with $|v| < \frac{1}{2}$. Finally, changing ivx to x, and adding, we obtain, in agreement with **(4: 19.7(11))**,

$$\frac{1 + 2vx + 4v^2}{(1+4v^2)^{3/2}}\, e^{4v^2x^2/(1+4v^2)} = \sum_{n=0}^{\infty} v^n H_n(x)/[\tfrac{1}{2}n]!$$

valid for all complex x, and all complex v with $|v| < \frac{1}{2}$.

References

1. G. Doetsch, *Handbuch der Laplace-Transformation* (Basel, 1950).
2. A. S. Eddington, *On a formula for correcting statistics for a known probable error of observation*, Monthly Notices Roy. Astron. Soc., *73* (1914), 359–360.
3. A. Erdélyi, et al., *Higher Transcendental Functions*, vol. 2 (New York, 1953).
4. ———, *Higher Transcendental Functions*, vol. 3 (New York, 1953).
5. I. I. Hirschmann and D. V. Widder, *The Convolution Transform* (Princeton, 1955).
6. H. Pollard, *Integral transforms*, Duke Math. J., *13* (1946), 307–330.
7. D. V. Widder, *The Laplace Transform* (Princeton, 1941).
8. ———, *Necessary and sufficient conditions for the representation of a function by a Weierstrass transform*, Trans. Amer. Math. Soc., *71* (1951), 430–439.

University of Toronto

INTERNATIONAL CONGRESS OF MATHEMATICIANS 1958

FIRST COMMUNICATION

At the invitation of the City and University of Edinburgh and the Royal Society of London, the International Congress of Mathematicians will meet in Edinburgh from August 14th to August 21st, 1958. His Royal Highness the Duke of Edinburgh has graciously consented to extend his patronage to the Congress.

The Executive Committee is inviting a number of mathematicians to deliver one-hour and half-hour addresses. There will also be daily sessions devoted to fifteen-minute communications. Those who wish to present such communications will have an opportunity of offering to do so when they receive the Second Communication.

There will be eight sections, namely:

1. Logic and Foundations.
2. Algebra and Theory of Numbers.
3. Analysis.
4. Topology.
5. Geometry.
6. Probability and Statistics.
7. Applied Mathematics, Mathematical Physics and Numerical Analysis.
8. History and Education.

A programme of entertainments and excursions is being planned.

There will be two categories of membership of the Congress:

Ordinary members, who will be entitled to participate in the scientific and social activities of the Congress and to receive the Proceedings of the Congress, and

Associate members accompanying ordinary members of the Congress, who will be entitled to many of the privileges of ordinary membership, but will not participate in the scientific programme and will not receive the Proceedings.

Those who wish to receive further information about the Congress are requested to communicate their name and full address to the Secretary as soon as possible. Towards the end of 1957 they will receive a Second Communication giving more detailed information and specifying the fees to be paid by members; a form of application for membership of the Congress will be included.

It would be of considerable assistance if, where possible, correspondents would use English, French, German or Russian.

Mathematical Institute,
16 Chambers Street,
Edinburgh, 1,
SCOTLAND.

FRANK SMITHIES,
Secretary of the Congress.
February 1957.

Royal
meet
hness
ge to

eliver
ed to
ations
econd

ic and
ngress,

, who
ill not
edings.

ss are
oon as
ication
mbers;

ndents

es,
gress.
57.